UNIVERSIDAD DE CANTABRIA

Programa de Doctorado en Ciencia y Tecnología

TESIS DOCTORAL

TÉCNICAS AUTOMATIZADAS DE DETECCIÓN Y ANÁLISIS DEL EFECTO DE LENTE GRAVITACIONAL EN CARTOGRAFIADOS DE GALAXIAS DE NUEVA GENERACIÓN

Alberto Manjón García

Directores: Dr. Jose M. Diego Rodríguez y Dr. Diego Herranz Muñoz

Escuela de Doctorado de la Universidad de Cantabria

Marzo 2021

Universidad de Cantabria
Departamento de Física Moderna

Instituto de Física de Cantabria
CSIC-Universidad de Cantabria

Automated techniques for the detection and analysis of the gravitational lensing effect in new generation galaxy surveys

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Physics by

Alberto Manjón García

Supervisors: Dr. Jose M. Diego Rodríguez and Dr. Diego Herranz Muñoz

Doctoral Program in Science and Technology

March 2021

Declaración de Autoría

Jose María Diego Rodríguez, Doctor en Ciencias Físicas e investigador científico del Consejo Superior de Investigaciones Científicas (CSIC),

y

Diego Herranz Muñoz, Doctor en Ciencias Físicas y profesor titular de la Universidad de Cantabria,

CERTIFICAN que la presente memoria titulada

Técnicas automatizadas de detección y análisis del efecto de lente gravitacional en cartografiados de galaxias de nueva generación

ha sido realizada por Alberto Manjón García bajo nuestra dirección en el Instituto de Física de Cantabria (IFCA), para optar al título de Doctor en Física por la Universidad de Cantabria. Consideramos que esta memoria contiene aportaciones científicas suficientemente relevantes como para constituir la Tesis Doctoral del interesado.

En Santander, a 10 de marzo de 2021

Jose María Diego Rodríguez

Diego Herranz Muñoz

A mis padres

(y a Chaqui y Gray)

You know, I used to live like
Robinson Crusoe – shipwrecked
among eight million people. Then
one day I saw a footprint in the
sand – and there you were.

C.C. "Bud" Baxter (The Apartment, 1960)

Agradecimientos

Intentaré ser breve, puesto que lo que trás esta página acecha no lo es. Este trabajo no habría sido posible sin la confianza, paciencia y ayuda que me habeís brindado, Diego y Chema. ¡Os estoy muy agradecido! Elena, Guillermo y Patricia, es un placer compartir despacho con vosotros. Es una lástima que la naturaleza nos haya arrebatado un año de esos momentos, de debates sobre política, y de pintar en la pizarra. Gracias, Patricio, por solucionarme dudas burocráticas y por darme el portátil con el que he realizado la mayor parte de lo aquí presentado. Gracias a todos con los que he pasado momentos durante estos años (Belén, Patxi, Enrique, Airam, Biuse, Raúl, Daniel y Marcos), bien haya sido en *coffee breaks*, en viajes por carretera, en avión, en visitas a colegios o en actividades de divulgación. Quiero mencionar también a todo los compañeros y personal del IFCA, especialmente a Silvia, que hacen de este edificio un magnífico lugar de trabajo. Por último, quiero expresar mi agradecimiento a todos los que han contribuido en una u otra forma al trabajo aquí reflejado. Principalmente a Helena y Jesús, con quiénes he aprendido mucho y pasado una grata estancia en Philadelphia. A Mariangela, por facilitar que mi estancia fuese posible. A Joaquín y Laura, cuyos comentarios ayudaron mucho a que la revisión de mi primer paper fuese muy liviana. A Daniel Lam, Andrea Biviano y Ben Metcalf por proporcionarme datos y resultados. Y para terminar, a todas aquellas personas cuyo trabajo ha hecho posible que yo haga el mio.

Contents

List of Figures

List of Tables

Introduction

The second half of the XXth century saw the flowering of observational cosmology. The detection of the cosmic microwave background (CMB) in 1964, along with the discovery of its structure and anisotropies in 1992; the observed abundances of light elements (hydrogen, deuterium, helium, and lithium) and the accelerating expansion of the Universe, measured and confirmed in 1998 with distant type Ia supernovae (SNe); and its large-scale structure constitute the outstanding observational pieces of evidence that support the ΛCDM model of cosmology. This parametric model, often referred to as the cosmological standard model, asserts that three main components exist in the Universe: first, the dark energy, which seems to behave like a cosmological constant (Λ) throughout space; second, the cold dark matter; and third, the ordinary matter. According to the last measurements of the CMB anisotropies (Planck Collaboration, 2020), which combine information from temperature and polarization maps with the reconstruction of the gravitational lensing effect suffered by the CMB signal, only 4.9% of the energy density in the Universe belongs to baryonic matter. The so-called dark matter, non-baryonic weakly interacting matter, represents 26.4%, while the remaning 68.7% of the energy density in the Universe is identified as dark energy.

Dark energy is responsible for the current accelerated expansion of the Universe (Riess et al., 1998; Perlmutter et al., 1999). Regardless of the fact that the dark energy equation of state has proven to be consistent with the cosmological constant Λ, and that the ΛCDM model fits the data well, this constant is still a phenomenological parameter, without any underlying theoretical explanation for its value (Weinberg, 1987). In addition, the empirically needed value of Λ marks our age as a special time in the evolution of the Universe. The Universe did not always expand at the rate it does today for the mere reason that its composition has changed throughout its history. Although now it is dominated by dark energy, the early Universe was dominated by radiation (photons and neutrinos), and so was the expansion. This radiation-dominated era lasted from the Big Bang until $z \sim 3300$. Then after this period of about 47,000 years, the Universe cooled enough so the density of matter (both ordinary and dark) outreached the density of radiation,

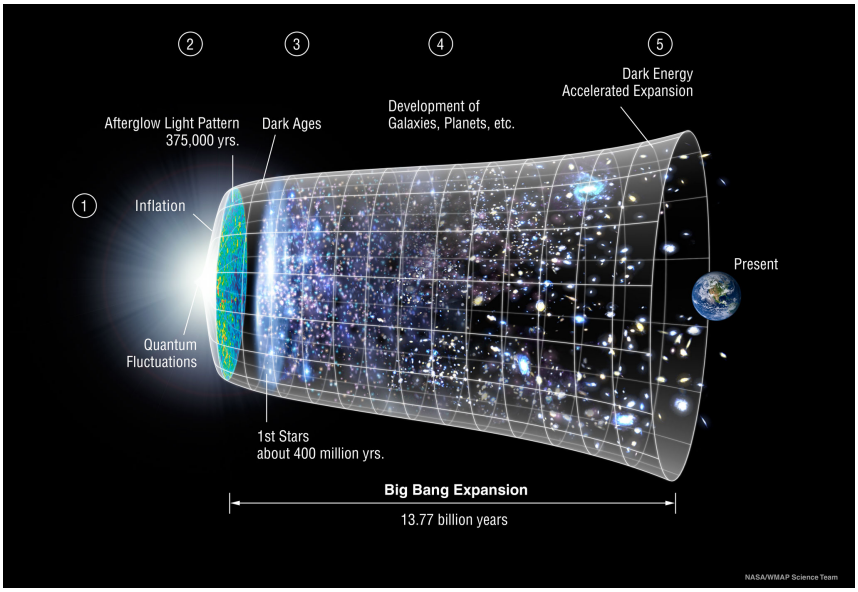


Figure 0.1: The evolution of the Universe as it is seen by the ΛCDM model. Courtesy of NASA/WMAP Science team.

and became dominant. During this matter-dominated era, which extended to $z \sim 0.3$, the expansion of the Universe began to slow down, as gravity countered the outward expansion. However, since the Universe continued to expand, the matter density kept on decreasing. Eventually, there came a point where the matter density dropped to such a low value that the contribution of another form of energy, the already mentioned dark energy, became significant enough, starting the dark-energy-dominated era we live in. This moment occurred when the Universe was about 9.8 billion years old, and since then the expansion rate of space began again to accelerate outwards, under the influence of the dark energy. As time goes on, the matter density will continue to drop, while the dark energy density will remain constant, which means that dark energy will become more and more dominant (see Huterer and Shafer, 2017, for a full review).

The accelerated expansion is one of the most mysterious phenomena in the Universe, but this is just one of the several consequences of the few we know about dark energy and dark matter. And thus, we are still far from having a deep enough knowledge of our Universe. For this reason, several attempts are being made in developing theories and models that agree with all the evidence that supports the standard model of cosmology but shed light where it nowadays does not. Disclosing the properties of this two dark components of the Universe is one of the challenges of cosmology for the years to come. This defiant task requires to be able to probe the early Universe in great detail. In this context, the development and testing of detection techniques of weak astronomical objects

at high redshift becomes a matter of great importance. In this thesis we will focus on the use of filtering techniques and on the exploitation of gravitation lensing to advance in our knowledge of the distant Universe.

Multifrequency detection of compact extragalactic sources

Over the last few decades, advances in the sensitivity of observations (especially in the infrared part of the spectrum) and progress in data processing have allowed us to probe the high redshift Universe in greater depth. The direct observation of galaxies in the redshift range $z \sim 1 - 10$ gives us the opportunity to study the cosmic history of galaxy and star formation at different cosmic epochs (see for example de Zotti et al., 2010; Eales, 2015). However, despite the constant increase in sensitivity of the detectors and in diameter of the telescopes, observations of the distant Universe are still flux-limited, rendering only those objects that are above the detection threshold of the instrument. In a Universe in which the inverse-square law prevails, a flux limit implies that the highest redshift galaxies accessible to any observatory will be among its faintest detectable objects. This situation is relieved for sources selected in the submillimetric range of the electromagnetic spectrum thanks to the strong, negative K correction, which leads to high-redshift galaxies being relatively easy to detect at submm wavelengths as compared with their low-redshift counterparts (Blain and Longair, 1993). In addition, lucky alignments of background objects with foreground lenses can push the limits further by enhancing the flux of objects that could not be detected otherwise. But even with the aid of the negative K correction and gravitational lensing, signal processing techniques are a fundamental tool to reach the faintest and most distant galaxies. This is particularly true in the microwave and far infrared (IR) parts of the electromagnetic spectrum, where the fluctuations from the cosmic infrared background (CIB) create a confusion noise whose level is comparable to the flux density of the typical high redshift galaxies.

The standard single-frequency detection methods for point sources in the CMB and far IR are based on wavelet techniques (Vielva et al., 2003; Barnard et al., 2004; González-Nuevo et al., 2006) or on the matched filter (or MF hereafter, Tegmark and de Oliveira-Costa, 1998; Herranz et al., 2002; Barreiro et al., 2003; López-Caniego et al., 2006, see also Herranz and Vielva (2010) for a review). Wavelets are well suited for the detection of compact sources due to their good position-scale determination properties, whereas the MF is the optimal linear detector-estimator because it provides the maximum signal-to-noise amplification for a source with a known shape (usually the point-spread function, or PSF hereafter, of the telescope) embedded in statistically homogeneous and spatially correlated

noise. By default, these techniques are only applicable to single-frequency sky images: even for multiwavelength observatories such as the *Herschel* Space Observatory (Pilbratt et al., 2010) or *Planck* (Tauber et al., 2010), the standard detection pipelines have produced individual source catalogs for each frequency band (see e.g., Planck Collaboration, 2011, 2014, 2016; Maddox et al., 2018). The next logical step is to boost the signal of faint sources by combining the different bands into a single detection, that is, "multifrequency detection". Most of the blind component separation algorithms that are used for diffuse components in microwave and far IR astronomy cannot deal with the high diversity of spectral behaviors associated to the different populations of extragalactic compact sources (see for example Leach et al., 2008). However, over the last few years a number of multifrequency compact source detection techniques have been proposed in the literature (Herranz and Sanz, 2008; Herranz et al., 2009; Lanz et al., 2010, 2013; Planck Collaboration, 2018). A review on the topic can be found in Herranz et al. (2012). In particular, if the spatial profile and the spectral energy distribution (SED) of the sources are known, and if the cross-power spectrum is known, or can be estimated from the data, the optimal linear detection method is the *matched multifilter* (or MMF hereafter, Herranz et al., 2002). This generalization outperforms the single-frequency MF in terms of signal-to-noise ratio.

The first of the works presented in this thesis consists of applying a *matched multifilter* to the first and second data releases of the *Herschel* Astrophysical Terahertz Large Area Survey (the *Herschel*-ATLAS or H-ATLAS, Eales et al., 2010), the largest single key project carried out in open time with the *Herschel* Space Observatory. We restrict our multifrequency analysis to the three wavelength bands covered by the SPIRE instrument aboard *Herschel* (Griffin et al., 2010), centered around 250, 350 and 500 μm. With this work we seeked to enhance the significance of detection of high-redshift extragalactic point sources in H-ATLAS, as well as to provide a sample of gravitational lens candidates.

Dark matter distribution in galaxy clusters

Dark matter constitutes one of the unknown components in the Universe. Since the mid-30s, it has been known that a meaningful fraction of the mass in the Universe is not tied up to the luminous stars, that once were thought to comprise most of the mass of galaxies (Zwicky, 1933). Observations of several galactic rotation curves (i.e., velocity distributions of stars and gas as they orbit the galactic center versus the distance to the center) lead to the conclusion that these velocity distributions cannot be explained theoretically only taking into account ordinary matter. The most straightforward explanation is that the mass of a galaxy has a significant non-luminous component, dark matter, and that the

visible matter represents just a small piece of the galaxy.

Even though it has not been directly detected yet, other indirect observations such as its influence on the large-scale structure of the Universe, its effects on the CMB and the temperature distribution of hot gas within galaxies and clusters, as well as the gravitational lensing effect by clusters of galaxies, also point to the presence of dark matter in the Universe. The only source of information to date for dark matter is due to its gravitational interaction with large concentrations of visible matter, such as stars or galaxies. If there are other types of interactions between visible and dark matter, they must be awfully weak as they have not yet been directly detected through experiments on Earth. Moreover, according to CMB spectrum, dark matter does not seem to self-interact or annihilates much either, supposing it does.

Although most of the matter in the Universe is dark matter, its underlying particle nature remains undisclosed. It is not even known if it consists of only one type of particle. What we know is that it cannot be made of ordinary particles, since otherwise the expected outcome from the early Universe nucleosynthesis would dissent with the current observed data. It does not absorb, reflect or emit any kind of electromagnetic radiation. Since dark matter forms part of the large structures we observe nowadays, such as galaxies and clusters, it must be mostly cold, in the sense of being a low-energy or non-relativistic particle. If dark matter was hot or warm, its presence in these gravitationally collapsed structures would be less noticeable than observed. Pursuant to most popular dark matter models, it must be very stable, i.e., if it decays, it must do so very slowly. This is owing to the fact that it is known that the proportion of dark matter in the early Universe was about the same as it is today. In this scenario, the mean lifetime of a dark matter particle would be at least 13.8 billion years.

The most broadly accepted hypothesis for dark matter nature is that it is composed of weakly interacting massive particles (WIMPs) that interact only through the weak force and gravity. There are other candidates for dark matter that include ordinary and heavy neutrinos, other postulated elementary particles like axions, light supersymmetric particles such as the neutralino, or massive astronomical bodies difficult to detect known as MACHOs (Massive Astrophysical Compact Halo Objects). There are also models built upon the assumption that the gravitation theory is wrong and they look for other alternatives so as to explain dark matter.

As it was mentioned above, the mass of the galaxies is dominated by dark matter, and the most revealing evidence we have for its existence has its origin in the gravitational interaction. Therefore, one of the ways of studying dark matter properties is to examine this interaction wherever the concentration of dark matter is large enough. Galaxy clusters seem

to be an excellent choice. And, indeed they are (see Kravtsov and Borgani, 2012). They are the most massive systems in dynamical equilibrium, and thus are highly sensitive and useful cosmological probes of gravitational structure formation and galaxy evolution throughout the history of the Universe. Given that galaxy clusters are believed to be dominated by dark matter, the measurement of their internal mass distribution can be helpful to constrain the physical properties of dark matter (Arabadjis et al., 2002; Markevitch et al., 2004; Serra and Domínguez Romero, 2011), and to distinguish between dark matter and alternative gravity theories (Clowe et al., 2006).

In this thesis we constrain the inner dark matter distribution of galaxy cluster MACS J1206.2-0847 through an analysis based on strong lensing data. Strong lensing is a gravitational lensing regime in which the light from a distant source, such as a quasar or galaxy, is deflected by a foreground galaxy or cluster of galaxies, resulting in multiple images of the background source. These images are also usually heavily distorted, acquiring the shape of rings or arcs. Studying these images provides us with unique information about how dark matter is distributed in galaxies and clusters, and helps us set limits on its self-interaction capacity. Independent measures of cosmological parameters, like the Hubble constant, have been made by measuring time delays in strong lensing events. The magnification effect of strong lensing, as can be appreciated in Fig. 0.2, enables us to observe and study faint and distant objects that otherwise would be inaccessible to observations. Strong lensing is, therefore, a major cosmological tool.

Figure 0.2: Hubble Space Telescope image of GAL–CLUS-022058s, the largest and one of the most complete Einstein rings discovered. It is nicknamed by astronomers as the 'Molten ring' due to its appearance. Credit to ESA/Hubble & NASA, S. Jha and L. Shatz.

Automatic identification of strong lenses

Gravitational lensing effect analyses would not be feasible if we could not detect it. The amount of data to be analyzed in astronomy is growing at a rate hard to follow with traditional approaches. World-class missions like Euclid, DES, SKA, LSST or J-PAS will provide images covering huge patches of the sky containing billions of galaxies. These large samples will supply us with more examples of rare events taking place in the Universe, from which we will be able to learn more. Strong gravitational lenses are among these rare events. Traditional search of lens features (arcs and rings) has relied exclusively on the visual inspection of the targets. However, this is an extremely time-consuming task for the current and future large-scale imaging surveys, given the much larger number of targets to be inspected. Therefore, finding strong gravitational lenses in this manner is difficult.

The development of automatic lens finder techniques has acquired great importance in recent years. Machine learning methods have been proven highly successful and can be used for a great variety of scientific purposes in extragalactic astronomy. We have found in Convolutional Neural Networks (CNNs) a good starting point to automatically recognize arc and ring patterns in images, and thus identify strong lensing systems. The last part of this thesis is focused on developing CNN-based models capable of identifying strong gravitational lenses in astronomical images from different kinds of surveys, with the highest possible precision. This work is still unpublished. In first place, we considered mock galaxy-galaxy lensing data, mimicking a ground-based, multi-band survey, roughly modeled on the Kilo-Degree Survey (KiDS). We have also trained a CNN with lensing simulations inspired in the observations expected from the Euclid satellite, both using only a single visible band and four different frequencies (VIS, J, Y, and H bands). Finally, we created galaxy-quasar strong lensing simulations based on the Javalambre Physics of the Accelerating Universe Astrophysical Survey (J-PAS). J-PAS is a unique survey with 54 narrow band filters, that will survey over 8000 square degrees and observe, among other objects, over half a million quasars, some of them strongly lensed. For our simulations we used mock quasars based on the J-PAS photometric system, along with galaxies, stars and quasars observed in a ~ 1 deg^2 area already explored.

The outline of this thesis is as follows. An overview of filtering aimed at the detection of point sources with the matched multifilter is shown in Chapter 1. Chapter 2 addresses the implementation of the MMF over H-ATLAS and details the results obtained (as it was published in Manjón-García et al., 2019). The basic theory of gravitational lensing required to understand the content of this thesis is depicted in Chapter 3. In Chapter 4 we present the free-form strong lensing analysis of galaxy cluster MACS J1206.2–0847 (as it was shown in Manjón-García et al., 2020). The theory behind the operation of artificial neural networks appears in Chapter 5. Chapter 6 displays the results obtained with our CNN in searching strong lenses using Euclid-like and KiDS-like mock lensing data. The preliminary performance of the CNN in searching lensed quasars using real and mock data based on J-PAS is shown in Chapter 7. Finally, in Chapter 8 we present a brief review of the conclusions reached in this thesis aimed to state its relevance for future work. A summary of this thesis written in Spanish is included in the last chapter.

Chapter 1

Filtering and detection methods

Filtering is a class of signal processing tool that allows to select the desired frequencies, or remove the unwanted ones, from raw data in order to favour the detection of a specific signal. From the mathematical point of view, a filter corresponds to the following operator:

$$L : f(x) \rightarrow g(x) = Lf(x) \tag{1.1}$$

where f is the input signal, L represents the filter, g is the output signal obtained after applying the filter, and x is the independent variable. A linear filter is one in which the filtered signal is a linear function of the input signal. The filter is homogeneous, or invariant in time, if the output is delayed by certain time τ when the input signal is also delayed by the same time. The homogeneity of a filter can be expressed as $g(x-\tau) = Lf(x-\tau)$. Most of the filters used in a wide range of scientific fields are linear and homogeneous. Any linear, time-invariant filter is completely characterized by its impulse response. This means that for any input, the output can be derived in terms of the input and the impulse response. The impulse response function of a filter is its output when a brief input signal, called impulse, is received. This impulse is usually modeled for continuous signals as a Dirac-δ distribution, whose value at x is obtained by the integral:

$$f(x) = \int f(u)\delta(x-u)du \tag{1.2}$$

The linearity and continuity of the filter L imply that

$$Lf(x) = \int f(u)L\delta(x-u)du \tag{1.3}$$

We use $h(x) = L\delta(x)$ to denote the impulse response of the filter L, so Eq. (1.3) can be written as follows:

$$Lf(x) = \int_{-\infty}^{\infty} f(u)h(x-u)du = \int_{-\infty}^{\infty} h(u)f(x-u)du = h \otimes f \tag{1.4}$$

where $\otimes$ stands for a convolution. This equation shows that a linear homogeneous filter equates to a convolution with the filter's impulse response h. The convolution theorem establishes that the convolution of two functions, h and f, at a coordinate point x can be expressed as:

$$Lf(x) = h(x) \otimes f(x) = \int_{-\infty}^{\infty} \hat{h}(q)\hat{f}(q)e^{-iqx}dq \tag{1.5}$$

where $\hat{h}(q)$ and $\hat{f}(q)$ are the corresponding Fourier transforms of $h(x)$ and $f(x)$, respectively, according to the following convention of the Fourier transform:

$$\begin{aligned} \hat{f}(q) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{iqx}dx \\ f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(q)e^{-iqx}dq \end{aligned} \tag{1.6}$$

The Fourier transform $\hat{h}(q)$ of the impulse response $h(x)$ is also known as the transfer function of the filter L. These considerations are valid for continuous signals. All in all, we see that the process of filtering an image with a linear homogeneous filter is equivalent to multiply the Fourier transform of the image with this transfer function. This transfer function can be considered then as a frequency selective device, in the sense of Fourier modes. This property makes working in the Fourier domain the best choice. In addition, the signals we want to detect belong to compact or point sources, whose profiles are well known. The Fourier modes of these compact sources are easily obtained from their profiles. For these reasons, filtering has proven to be very well suited for the detection of compact sources embedded in a noisy background. We only have to design transfer functions able to reduce the contribution of the frequencies responsible for the noise while preserving the characteristic frequencies corresponding to the compact sources. The widely used band-pass filters, where the filter is set to zero outside the range of frequencies we want to preserve, are built on the basis of this idea:

$$\hat{h}(q) = \begin{cases} 1 & \text{if } |q| < q_c \\ 0 & \text{if } |q| \geq q_c \end{cases}$$

where q_c is the cutoff frequency. The region $|q| < q_c$ is the pass-band of the filter, while the region $|q| \geq q_c$ is the stop-band. The impulse response of a band-pass filter is:

$$h(x) = 2q_c \frac{\sin(q_c x)}{x} \tag{1.7}$$

Other related filters are the *low-pass* and the *high-pass* filters. Low-pass filters, which remove the higher frequencies, are useful to smooth images since the small scale details come from high frequencies in the Fourier domain. On the contrary, high-pass filters, which remove the low frequencies, are used to mitigate the effect of large-scale contaminants. However, the discontinuous shape of the transfer function in these filters can introduce ring-shaped artifacts in the image in the real space. After the filtering process, this effect gives rise to spurious compact objects, which can be mistaken as real objects. Despite the fact that band-pass filters are used in many disciplines of signal processing, they are not suited for the detection of compact sources in astronomical images due to this reason. The kind of filter proper for filtering astronomical images must start by being continuous in the Fourier space, in order to avoid this type of contamination. The next step is to find a filter that improves the detection of compact sources embedded in a background by reducing the background noise and preserving the signal. The most commonly used criterion to achieve this seeks to increase as much as possible the signal-to-noise ratio (S/N) of the sources we are trying to detect. Let us assume a signal s with amplitude A at the position x_0 inlaid in a noisy background n with dispersion σ. The S/N of this signal is defined as the ratio between its amplitude and the standard deviation of the whole field:

$$\mathrm{S/N} = \frac{s(x_0)}{\sigma} = \frac{A}{\sigma} \tag{1.8}$$

Obviously, the S/N of the field will change depending on the filter considered. Therefore, it is essential to be able to measure the performance of any filter used. The gain or amplification λ of the signal achieved with a given filter is defined as:

$$\lambda = \frac{\mathrm{S/N}_\psi}{S/N} = \frac{s_\psi(x_0)/\sigma_\psi}{s(x_0)/\sigma} = \frac{s_\psi(x_0)\sigma}{s(x_0)\sigma_\psi} \tag{1.9}$$

where $s_\psi(x_0) = \psi \otimes s(x_0)$ is the filtered map at the position of the source and σ_ψ is the

standard deviation of the filtered map. If the amplification is greater than one, the chances to detect the sought signal increase. In this sense, the optimal filter for the detection of compact sources would be the one that maximizes Eq. (1.9).

We introduce in the following pages the matched filter, one of the standard single-frequency detection methods for compact sources. A single-frequency technique is only able to filter images taken at different frequencies separately. Therefore, the filtered maps must be combined a posteriori in order to perform a multifrequency-like analysis. Afterwards, we approach the multifrequency methods, starting with the matched matrix filters, and ending with the matched multifilter explored in this work.

1.1 Matched filter

The way of maximizing the gain λ of a filter, defined in Eq. (1.9), is to minimize σ_ψ while preserving the signal of interest, $s_\psi(x_0) = s(x_0)$. The filter that has proven to be able to maximize the S/N of a signal in a single frequency channel, and therefore can provide the highest possible gain, is the *matched filter* (MF). In the following lines, we are deriving the expression of the two-dimensional MF. Although this is a single-frequency method, here we will already consider the existence of observations at various frequencies. We do this in order to lay the foundation for the later introduction of the *matched multifilter*, which is based on the MF method.

First of all, let us assume that several images at N different frequencies are taken of the same area of the sky. For simplicity, we also assume that there is only one point source, of amplitude A_ν at each frequency, centered at the coordinate origin of these images. The signal in these images can be described as:

$$y_\nu(\mathbf{x}) = f_\nu s_\nu(\mathbf{x}) + n_\nu(\mathbf{x}) \tag{1.10}$$

where y_ν is the total signal in the pixel $\mathbf{x}$, f_ν is the frequency dependence of the point source, s_ν is the contribution of the point source to the total signal, and n_ν represents the background noise in each pixel, containing both the instrumental noise as the one belonging to other contributions detected. These variables are measured at every observing frequency $\nu = 1, ..., N$ considered.

Before getting any image, the signal from this point source reaches a detector. The angular resolution of this detector is greater than the intrinsic angular size of such point source. Its signal is then convolved, at each observing frequency ν, with the corresponding

antenna beam. We assume, again for simplicity, that the antenna beams can be well described by symmetric 2D Gaussian functions. The signal from the point source can then be written as:

$$s_\nu(x) = A_\nu \tau_\nu(x) \tag{1.11}$$

where $\tau_\nu(x)$ is the spatial profile of the source, and $x = |\mathbf{x}|$ due to the assumption that the beams considered are symmetric. The background noise $n_\nu(\mathbf{x})$ can be modeled as a homogeneous and isotropic random field with average value equal to zero and whose power spectrum P_ν is defined by:

$$\langle n_\nu(\mathbf{q}) n_\nu^*(\mathbf{q'}) \rangle = P_\nu(\mathbf{q}) \delta_D^2(\mathbf{q} - \mathbf{q'}) \tag{1.12}$$

where $n_\nu(\mathbf{q})$ is the Fourier transform of $n_\nu(\mathbf{x})$, δ_D^2 is the 2D Dirac distribution, and the symbol $*$ denotes complex conjugation.

Hereinafter, we will focus solely on the signal from one of the ν frequencies observed. According to the convention of the Fourier transform chosen in Eq. (1.6) and the convolution theorem declared in Eq. (1.5), if we decide to apply an unknown filter ψ on the image signal from Eq. (1.10), the resulting filtered field at a certain coordinate point $\mathbf{b}$ is described by:

$$\omega_\psi(\mathbf{b}) = A\omega_\tau(b) + \omega_n(\mathbf{b}) = \int y(\mathbf{q}) \psi^*(\mathbf{q}) e^{-i\mathbf{q}\cdot\mathbf{b}} d\mathbf{q} \tag{1.13}$$

where

$$\omega_\tau(\mathbf{b}) = \int \tau(\mathbf{q}) \psi^*(\mathbf{q}) e^{-i\mathbf{q}\cdot\mathbf{b}} d\mathbf{q} \tag{1.14}$$

$$\omega_n(\mathbf{b}) = \int n(\mathbf{q}) \psi^*(\mathbf{q}) e^{-i\mathbf{q}\cdot\mathbf{b}} d\mathbf{q} \tag{1.15}$$

Given that both the source profile and the filter are assumed to be circularly symmetric ($\psi(\mathbf{q}) = \psi(q)$), and since the only point source is placed at the coordinate origin, i.e., $\mathbf{b} = 0$, the filtered field at the position of the source is given by:

$$\omega_\psi(0) = \omega_\tau(0) = 2\pi \int_0^\infty \tau(q) \psi^*(q)\, q\, dq \tag{1.16}$$

And the variance of the filtered field is

$$\sigma_\psi^2 = \int_0^\infty P(\mathbf{q})|\psi|^2(\mathbf{q})\, d\mathbf{q} = 2\pi \int_0^\infty P(q)|\psi|^2(q)\, q\, dq \tag{1.17}$$

Now it only remains to find the filter ψ that provides a total filtered field ω_ψ optimal for the detection of the sources. This filter must meet the following two requirements:

1. The filter must be an *unbiased* estimator of the source amplitude. This means that, after filtering the image, the amplitude value A of the source has to be recovered, on average over many realizations, at the source position. In a specific case, even after filtering, there will always be some contribution of the noise.
2. The filter must be a *maximum efficiency* estimator. This means that the variance σ_ψ^2 around the average amplitude value A recovered has to be as small as possible.

The condition (1), written above, to be met by the filter can be expressed as:

$$\langle \omega_\psi(0) \rangle = A\langle \omega_\tau(0) \rangle + \langle \omega_n(0) \rangle = A \tag{1.18}$$

Since the noise has zero mean, $\langle n(\mathbf{x}) \rangle = 0$, it is also true that $\langle n(\mathbf{q}) \rangle = 0$, and therefore

$$\langle \omega_n(0) \rangle = \langle \int n(\mathbf{q})\psi^*(\mathbf{q}) d\mathbf{q} \rangle = \int \langle n(\mathbf{q}) \rangle \psi(\mathbf{q}) d\mathbf{q} = 0 \tag{1.19}$$

Consequently, Eq (1.18) is only satisfied if

$$2\pi \int_0^\infty \tau(q)\psi^*(q)\, q\, dq = 1 \tag{1.20}$$

Regarding the condition (2) of maximum efficiency, the variance σ_ψ^2 comes only from the noise component (n) because the signal component ($A\tau$) is always the same. Thus, we have to minimize Eq. (1.17) while satisfying Eq. (1.20). The way of minimizing a function with a ligation is through the method of Lagrange multipliers. We define a Lagrangian that depends on the still-unknown filter ψ:

$$\begin{aligned} \mathcal{L}(\psi) &= \sigma_\psi^2 + \lambda\Big[2\pi \int_0^\infty \tau(q)\psi^*(q)\, q\, dq - 1\Big] \\ &= 2\pi \int_0^\infty P(q)|\psi|^2(q)\, q\, dq + \lambda\Big[2\pi \int_0^\infty \tau(q)\psi^*(q)\, q\, dq - 1\Big] \end{aligned} \tag{1.21}$$

Taking variations with respect to ψ we get:

$$\mathcal{L}(\psi + \delta\psi) - \mathcal{L}(\psi) = 4\pi \int_0^\infty P(q)\, \psi^*\, \delta\psi^*\, q\, dq + 2\pi\lambda \int_0^\infty \tau(q)\delta\psi^*\, q\, dq \tag{1.22}$$

We can equal the result to zero in order to get:

$$2\pi \int_0^\infty [2P(q)\psi(q) + \lambda\tau(q)]\delta\psi^*\, q\, dq = 0 \tag{1.23}$$

A sufficient condition for this last equation to be satisfied is:

$$\psi(q) = \frac{\lambda\tau(q)}{2P(q)} \tag{1.24}$$

Taking into account that $\tau(x)$ and $\tau(q)$ are both real functions with radial symmetry, the value of λ can be obtained substituting the filter expression above in Eq. (1.20):

$$\lambda = \left(\pi \int_0^\infty \frac{\tau^2(q)}{P(q)}\, q\, dq\right)^{-1} \tag{1.25}$$

Finally, all this leads us to the formula of the *matched filter* in the Fourier space:

$$\psi(q) = \frac{1}{\alpha}\frac{\tau(q)}{P(q)} \tag{1.26}$$

where α is a normalization factor that preserves the source amplitude after the filtering process:

$$\alpha = 2\pi \int_0^\infty \frac{\tau^2(q)}{P(q)}\, q\, dq \tag{1.27}$$

The *matched filter* (Tegmark and de Oliveira-Costa, 1998; Barreiro et al., 2003; López-Caniego et al., 2006) is established in the literature as the standard single frequency method for the detection of point sources. This filter gives the maximum S/N in the presence of additive stochastic noise, which makes it the optimal linear detector for single frequency maps. In any single frequency approach, each frequency channel is processed separately and independently from the other channels. The robustness of this approach is that no a priori assumption about the spectral behavior of the sources considered is required in order to be successfully applied. However, its main drawback is the loss of the potential noise reduction that could be obtained by using the spectral information at different frequencies.

We want to end the explanation of the MF, recalling that the impulse response of any linear, time-invariant filter, such as the MF, can be characterized when a brief input signal is received. Let us consider then a Dirac-δ signal from a compact object in the sky. This signal is observed with a detector whose PSF is a Gaussian of width σ. Due to the intrinsic characteristics of the measurement procedure, after passing through the detector the signal becomes a Gaussian, centered at the position of the original one, but surrounded by white instrumental noise that makes it more difficult to be detected. Now, let us assume that the signal received is dominated by the noise and the contribution of the point source is minimal. Therefore, the power spectrum of the noise is a good approximation of the power spectrum $P(q)$ of the whole signal. As it is white noise, $P(q)$ is approximately constant. Under all these circumstances, the MF, given by Eq. (1.26), suitable to filter the signal will be proportional to the Gaussian of the PSF. This important result can be stated as the following: the optimal filter to denoise a signal with a Gaussian profile embedded in white noise is a Gaussian filter. Signals in scales lower than the σ of the Gaussian, responsible for the noise, are removed while signals at larger scales, belonging to compact sources, and whose shape is closer to the PSF of the detector, are preserved.

In a more realistic scenario than the one depicted above, the MF is not free of some drawbacks which deserve to be mentioned. In order to construct the MF properly, it is necessary to estimate the power spectrum for all the Fourier modes in the signal. Despite the fact that this power spectrum, shown in Eq. (1.12), can be approximated from the generalised noise, it is particularly difficult to carry it out for the low modes, which are pretty noisy. The use of these noisy modes to design the MF often results in a discontinuous filter in the Fourier space. These discontinuities turn into ring-shaped effects when the image is filtered. These effects can be mitigated by smoothing the spectrum before building up the filter, but we are inevitably manipulating the input image. Lastly, there will be times when some of the Fourier modes could not be estimated, and will therefore have to be guessed.

1.2 Matched matrix filters

The moment has come to address the multifrequency techniques. In a multifrequency approach, the frequency dependence of the sources and statistical correlation of the noise between the N different observing frequencies are taken into account. Unlike the single frequency approach, an important noise reduction can be achieved this way, provided that there is a noise correlation between the different maps. As stated earlier for the *matched filter* in Sec. 1.1, the signal $y_\nu(\mathbf{x})$ in N two-dimensional images can be described with

Eq. (1.10) and the background noise $n_\nu(\mathbf{x})$, in which there is an unknown number of point sources, can be modeled as a homogeneous and isotropic random field with average value equal to zero. The cross-power spectrum $\mathbf{P} = (P_{\nu_1\nu_2})$ between the noises at any two of these frequencies, ν_1 and ν_2, is defined by:

$$\langle n_{\nu_1}(\mathbf{q}) n^*_{\nu_2}(\mathbf{q'})\rangle = P_{\nu_1\nu_2}\delta^2_D(\mathbf{q}-\mathbf{q'}) \tag{1.28}$$

The *matched matrix filters* is a linear filtering technique (Herranz and Sanz, 2008; Herranz et al., 2009, MTXF) that maximizes the signal-to-interference ratio of compact sources embedded in a set of images by taking advantage of the background's power spectrum for each channel and its cross-correlation among the different channels. The MTXF does not make any a priori assumption about the spectral behavior of the sources. The other two basic underlying ideas of this method are that the same source will appear in the same position in all the images, and that although the spatial profile of the sources may differ from channel to channel, it is a priori known. In order to achieve an accurate photometry in all available channels, it is required a transformation in which the number of output maps processed is equal to that of the input images. And, as a multifrequency approach, all N channels have to contribute in the elaboration of each one of the output maps. The possibility proposed in Herranz and Sanz (2008) is to define a set of $N \times N$ filters such that the processed maps are described by:

$$\omega_{\nu_1}(\mathbf{x}) = \sum_{\nu_2} \int \Psi_{\nu_1\nu_2}(\mathbf{q}) y_{\nu_2}(\mathbf{q}) e^{-i\mathbf{q}\cdot\mathbf{x}} d\mathbf{q} \tag{1.29}$$

where $\mathbf{\Psi} = \Psi_{\nu_1\nu_2}$ is the matrix of filters. This matrix filter must satisfy the two optimality requirements, (1) and (2), for the detection of point sources detailed in Sec. 1.1. These two conditions ensure that the filtered maps ω_{ν_i} are both unbiased and efficient. The set of filters that minimize the variance $\sigma_{\omega_{\nu_i}}$ at each frequency ν_i while keeping the individual amplitudes A_{ν_i} constant for all point sources, independently of their frequency dependence, can be shown to be given by the following expression:

$$\mathbf{\Psi^*} = \mathbf{F}\mathbf{P}^{-1} \tag{1.30}$$

where $\mathbf{F} = (F_{\nu_1\nu_2}) = (\lambda_{\nu_1\nu_2}\tau_{\nu_2})$, being $\lambda = (\lambda_{\nu_1\nu_2}) = \mathbf{H}^{-1} = (H_{\nu_1\nu_2})^{-1}$, and

$$H_{\nu_1\nu_2} = \int \tau_{\nu_1}(\mathbf{q}) P^{-1}_{\nu_1\nu_2}(\mathbf{q}) \tau^*_{\nu_2}(\mathbf{q})\, d\mathbf{q} \tag{1.31}$$

It is pretty straightforward to see from Equations (1.30) and (1.31) that, for the case where the noise is completely uncorrelated between channels, the matrix filter $\mathbf{\Psi^*}$ becomes a diagonal matrix whose non-zero elements are the complex conjugates of the matched filters corresponding to each input channel. For sources with circularly symmetric profiles, the filters are real-valued and the whole process is equivalent fo filter each channel independently with the proper *matched filter* defined in Eq. (1.26). However, when there is noise correlation between channels, a substantial reduction of the background is achieved with the MTXF thanks to the optimal combination of the filtered maps. This technique has been tested on CMB simulations at different wavelengths outperforming (or at least equaling) the results obtained with the standard mono-frequency MF, introduced in Sec. 1.1, in terms of reliability, completeness and flux accuracy.

1.3 Matched multifilter

In Sec. 1.2 we said that the MTXF does not make any a priori assumption about the spectral behavior of the point sources. As a consequence, this filtering results in a $N \times N$ matrix of filters which generates a filtered map for each value of A_ν. When a priori assumption about the spectral behavior of the sources is made, A_ν consists of a single value (A), and the matrix can be reduced to a $N\times$ 1 vector of filters. The optimal filters for the detection of point sources belonging to this vector are what we know as the *matched multifilter* (MMF). This is the multifrequency method explored in this work. A detailed derivation of the MMF formula, which is strongly based on the derivation of the MF shown in Sec. 1.1, is presented in Appendix A. Under the same assumptions for a multifrequency approach, regarding the treatment of the signal and the noise, established in sections 1.1 and 1.2, the set of N linear time-invariant filters ψ_ν that made up the MMF are defined as

$$\omega_{\psi_\nu}(\mathbf{b}) = \int y_\nu(\mathbf{x})\psi_\nu(\mathbf{x};\mathbf{b})d\mathbf{x} = \int y_\nu(\mathbf{q})\psi_\nu(\mathbf{q})e^{-i\mathbf{q}\cdot\mathbf{b}}d\mathbf{q} \tag{1.32}$$

where $\mathbf{b}$ defines a translation from the coordinate origin, $\omega_\nu(\mathbf{b})$ represents each filtered image at frequency ν in the position $\mathbf{b}$, and $y_\nu(\mathbf{q})$ and $\psi_\nu(\mathbf{q})$ are the corresponding Fourier transforms of $y_\nu(\mathbf{x})$ and $\psi_\nu(\mathbf{x})$, respectively, according to the Fourier transform convention defined in Eq. (1.6). The first step in this multifrequency method is thus to filter each image y_ν with one of these linear filters ψ_ν. In the next step, all the filtered maps ω_{ψ_ν} from Eq. (1.32) are combined according to:

$$\omega_\psi(\mathbf{b}) = \sum_\nu \omega_{\psi_\nu}(\mathbf{b}) \tag{1.33}$$

This total filtered map is an image in which the signal $s(x)$ has been enhanced while the noise $n(x)$ has been significantly decreased. The two necessary requirements to ensure that the filtered field is optimal for the detection of point sources in this multifrequency approach are the same as those, (1) and (2), explained in Sec. 1.1. On the one hand, the total filtered map must be an unbiased estimator of the amplitude of the source, so the filters ψ_ν must be unbiased. On the other hand, the variance of the total filtered map ω_ψ has to be as small as possible, i.e., the filters must be maximum efficiency estimators.

As we said at the beginning of this section, the filter we are looking for is called *matched multifilter* (MMF) and was introduced for the first time for the detection of Sunyaev-Zel'dovich (SZ) clusters in multifrequency maps Herranz et al. (2002). The MMF is the optimal linear detection method when the frequency dependence and the spatial profile of the sources are known, or considered known, and the cross-power spectrum of the noise is known or can be estimated from the data. In the Fourier space the MMF can be written as follows:

$$\begin{aligned} \mathbf{\Psi}(q) &= \alpha\, \mathbf{P}^{-1}\mathbf{F}, \\ \alpha^{-1} &\equiv \int \mathbf{F}^t\, \mathbf{P}^{-1}\, \mathbf{F}\, dq, \\ \sigma^2 &= \int \mathbf{\Psi}^t\, \mathbf{P}\, \mathbf{\Psi}\, dq = \alpha \end{aligned} \tag{1.34}$$

where $\mathbf{\Psi}(q) = [\psi_\nu(q)]$ is the column vector of the filters; $\mathbf{F} = [f_\nu \tau_\nu]$, being f_ν the frequency dependence and τ_ν the source profile at each frequency ν; $\mathbf{P}^{-1}$ is the inverse matrix of the cross-power spectrum $\mathbf{P}$; and σ^2 is the variance of the output-filtered image. In Eq. (1.34) and in the following discussion, $q \equiv |\mathbf{q}|$ is the modulus of the Fourier wave vector since we are assuming circularly symmetric source profiles. And since the cross-power spectrum only depends on the modulus q, all the formulas can be expressed in terms of q instead of the full vector. However, it would be easy to generalize our formulas for non-symmetric profiles just by replacing q by $\mathbf{q}$ in the equations. Finally, α in Eq. (1.34) can be interpreted as the normalization that is requested in order to guarantee that the filters ψ_ν are unbiased estimators of the flux density of the sources under study. Further details can be found in Herranz et al. (2002); Lanz et al. (2010, 2013).

It can be easily seen the formal similarity between the MMF formula and the expression for the MF shown in Eq. (1.26). Both are directly proportional to the profile τ of the source, and inversely proportional to the power spectrum $\mathbf{P}$ of the noise. In fact, as we pointed out for the MTXF expression (see Eq. 1.30), the MMF reduces to a diagonal matrix of N *matched filters*, one at each frequency, if the maps considered do not have an underlying noise correlation between different frequencies. Therefore, a proper multifrequency approach can only be carried out as long as the maps considered show correlations among them. The spectral behavior f_ν of the sources is not known a priori just with the information of the images. Therefore, some way to model this unknown frequency dependence needs to be found. Lanz et al. (2010) showed that the MMF could be generalized for the case where this dependence is not known. Despite this, in this work we will incorporate a specific SED to the MMF in order to derive photometric redshift estimations of the extragalactic sources under study. In order to address this problem, it is very useful to rewrite the vector $\mathbf{F} = [f_\nu \tau_\nu]$ in the matrix form $\mathbf{F} = \mathbf{T}(q)\boldsymbol{f}(\nu)$, with diagonal matrix $\mathbf{T}(q) = [\tau_1(q),..., \tau_N(q)]$ and the vector of frequency dependence $\boldsymbol{f}(\nu) = [f_\nu]$. Thereby, we are able to include all the dependence of q in the matrix $\mathbf{T}$, pulling it completely apart from the dependence in ν. This way Eq. (1.34) can be rewritten as:

$$\boldsymbol{\Psi}_{\boldsymbol{f}}(q) = \alpha_f \mathbf{P}^{-1}\mathbf{T}\boldsymbol{f}, \quad \alpha_f^{-1} = \int \boldsymbol{f}^t \mathbf{T}\mathbf{P}^{-1}\mathbf{T}\boldsymbol{f} dq = \boldsymbol{f}^t \mathbf{H}\boldsymbol{f} \tag{1.35}$$

where matrix $\mathbf{H} = \int \mathbf{T}\mathbf{P}^{-1}\mathbf{T} dq$ and we based on the fact that $\mathbf{T}^t = \mathbf{T}$ and that vector $\boldsymbol{f}$ does not depend on q. This reformulation of the MMF is very convenient for its implementation. The most time-consuming part of the filtering is the calculation of the matrices $\mathbf{P}$ and $\mathbf{T}$ since they must be calculated for all values of q. In the case we are considering in this work the only quantity that varies during the maximization process is the redshift of the source we want to estimate. This allows us to compute the integrals of matrix $\mathbf{H}$ only once for each set of images of the source considered. The estimation of the amplitude A of a point source present in a set of images after applying these filters is given by:

$$A_{\Psi_f} = \omega_{\Psi_f}(0) = \alpha_f A \boldsymbol{f}^t \mathbf{H}\boldsymbol{f} \tag{1.36}$$

Chapter 2

Application of the MMF over H-ATLAS sources

In this chapter, we apply the *matched multifilter* technique to the multi-band infrared astronomical images of the *Herschel*-ATLAS survey, aiming at producing a catalog of distant and faint IR sources. Despite the fact that Lanz et al. (2010) showed that the MMF can be generalized for the case where the SED of the sources is not known, we will incorporate here a specific SED so as to derive photometric redshift estimations of high-redshift dusty star-forming galaxies (DSFGs) detected in the IR part of the spectrum[1]. We restrict our multifrequency analysis to the three wavelength bands covered by the SPIRE instrument aboard *Herschel* (Griffin et al., 2010), centered around 250, 350 and 500 μm. As discussed in Negrello et al. (2010), Lapi et al. (2011), González-Nuevo et al. (2012), Pearson et al. (2013) and Donevski et al. (2018), the SPIRE bands are ideal for capturing the peak in the SED corresponding to the dust emission of SFGs at $z \sim 2$. The dust absorbs the light from their stars and reemits it at longer wavelengths. Therefore, their emission is redshifted from its rest-frame wavelength around 70–100 μm to the SPIRE wavelengths. Although galaxies have formed most of their stars at this redshift range, at higher redshifts DSFGs occupy the most massive halos and are among the most luminous objects found at $z \sim 4$ (Michałowski et al., 2014; Oteo et al., 2016; Ikarashi et al., 2017). These high-redshift DSFGs have markedly red colors as seen by SPIRE, with rising flux densities from 250 to 500 μm (the so-called "500 μm-risers"), and have received a great deal of attention in the recent years (see for example Ivison et al., 2016; Negrello et al., 2017; Strandet et al., 2017). The DSFGs, and particularly the 500 μm risers uncovered by *Herschel*, are providing much insight into the early star forming history of the Universe.

[1] As measured at the observer's rest frame.

However, limited angular resolution and sensitivity severely constrain the power of this type of objects as astrophysical probes. The sensitivity of SPIRE allows for the direct detection of only the brightest, and thus rarest objects, at the bright end of the luminosity function (Negrello et al., 2017). By means of our multifrequency MMF technique, we intend to enhance the detectability and statistical significance of very faint red objects in the H-ATLAS source catalog and so expand the list of reliable 500 μm-riser candidates.

From the catalog of distant and faint sources derived, we also select a sample with those that are more likely to be gravitationally lensed. Although a non-negligible part of the faint H-ATLAS sources at $z > 1$ could be detected thanks to having been amplified by weak lensing (González-Nuevo et al., 2014, 2017), most of the faint high-z candidates in the H-ATLAS catalog have not been strongly lensed (with magnification factors larger than a few) by foreground halos (Negrello et al., 2017). In the other end of the flux density distribution, gravitational lensing plays an important role by magnifying distant galaxies that could be otherwise below the detection threshold or, at the very least, be observed with a significantly smaller flux (González-Nuevo et al., 2012; Bussmann et al., 2012, 2013; Messias et al., 2014). Gravitational lensing is a powerful astrophysical and cosmological probe, particularly rewarding at submillimeter wavelengths (Negrello et al., 2007; Vieira et al., 2013). As mentioned earlier, submillimeter telescopes such as *Herschel* have limited spatial resolution and consequently high source confusion, which makes probing the DSFGs difficult. By increasing the apparent angular size and brightness of the sources, gravitational lensing helps to detect and carry out follow-up observations of distant galaxies obscured by dust (Cox et al., 2011; Conley et al., 2011; Negrello et al., 2014; Nayyeri et al., 2016; Spilker et al., 2016). Even though, the identification of gravitational lenses is tricky and usually results in few candidates due to the relatively low probability of lensing, wide-area submillimeter surveys can simply, and easily detect strong gravitational lensing events at high fluxes, with nearly full efficiency, as was proved by Negrello et al. (2010). These are often strongly lensed galaxies (SLGs) with magnification factors of order ten that can be more easily detected owing to their magnified flux (Wardlow et al., 2013). The identification of these lenses is of great interest for multiple reasons. They offer the possibility to study in greater detail distant galaxies and resolve some of their features (Canalog et al., 2014; Dye et al., 2015). Also, the background galaxies can be used to reveal the internal structure of the lenses. Having a large catalog of SLGs will be important in future studies. For example, caustic crossing events on these galaxies can be used to study, not only distant luminous stars, but also the constituents of the lens itself. If a sizable fraction of dark matter is made of compact objects, caustic crossing events can be used to set limits in their fraction on a range of masses from subsolar to tens of solar masses. This mass range can be difficult to probe otherwise (see for instance Diego et al., 2018b).

2.1 H-ATLAS

Herschel-ATLAS is an extragalactic survey covering the widest area undertaken with *Herschel* Space Observatory (Pilbratt et al., 2010), imaging 659.25 deg^2 of the sky distributed in five fields: three (GAMA9 with 53.43 deg^2, GAMA12 with 53.56 deg^2 and GAMA15 with 54.56 deg^2) on the celestial equator, a large field (180.1 deg^2) centered on the north Galactic pole (NGP), and an even larger field (317.6 deg^2) centered on the south Galactic pole (SGP). Images have been taken in five far-IR to sub-mm photometric bands using the PACS (100 and 160 μm) and SPIRE (250, 350 and 500 μm) instruments in parallel mode. PACS measurements have not been used in the analysis performed in this work. The main reason is that the SED model assumed to estimate the redshifts of the H-ATLAS sources, see Eq. (2.1), has been developed to use only the SPIRE fluxes. This was owing to the fact that not all H-ATLAS sources have flux measurements at PACS wavelengths and only a few per cent of them were detected at greater than 5σ in these bands. However, SPIRE bands themselves are ideal for capturing the emission peak belonging to the high-redshift sources aimed in this work. The maps at different bands produced from SPIRE are close in terms of frequency and show noise correlation between them, allowing to perform a multifrequency approach.

Both Data Release 1 (DR1) and the subsequently released Data Release 2 (DR2) have been used in this analysis. *Herschel*-ATLAS DR1 includes the three equatorial fields covered by the Galaxy And Mass Assembly (GAMA) (Driver et al., 2009, 2016) spectroscopic survey. The three fields are $\sim$162 deg^2 combined, and are approximately located around 9h, 12h and 15h in α. The associated catalog, described in Valiante et al. (2016) and Bourne et al. (2016), covers all three regions and includes 120,230 SPIRE sources, which have at least a S/N = 4σ (including confusion noise) in any of the 250, 350 or 500 μm maps. *Herschel*-ATLAS DR2 covers the two fields centered in the NGP and SGP, which are about 450 deg^2 combined. The maps are described in Smith et al. (2017) while the submillimeter catalog is described in Maddox et al. (2018)[2] and include 118,980 (NGP) and 193,527 (SGP) sources, respectively. These sources also have at least a S/N = 4σ detection in all of the SPIRE bands. The complete H-ATLAS catalog contains a total of 432,737 sources, most of them being point sources. The resulting catalog after removing extended sources and stars still has 410,997 sources.

As it is explained in greater detail in Valiante et al. (2016), sources were detected using the MADX algorithm (Multiband Algorithm for source Detection and eXtraction) applied to the SPIRE maps. The first step of this method is to use *Nebuliser* to remove the

[2]H-ATLAS catalogs are public and available at http://www.h-atlas.org/public-data/download.

diffuse Galactic dust emission from all maps in the three bands, resulting in raw images with the local large-scale background subtracted ("backsub" maps). Then the images are convolved with a proper matched-filter for each band (Chapin et al., 2011). Maps of the variance in each of these convolved maps are also created. During the convolution, the contribution of each pixel of the input image is weighted by the inverse of the square of the instrumental noise in that pixel. The resulting maps are background subtracted maps and noise-weighted maps filtered with a customized matched filter ("fbacksub" maps). In the next step (in MADX), the maps at 350 and 500 μm are interpolated onto images with the same pixel scale as the 250 μm one, and the three images and their corresponding variance maps are then combined together to form a single signal-to-noise or *detection* image. In practice, images at 350 and 500 μm are given a zero weighting regarding source detection, that is, the detection image in MADX is simply the 250 μm image. The position of the source in this image will be used to estimate the fluxes of the source in the 350 and 500 μm maps. A list of potential sources is produced by finding all peaks in the detection image with S/N $> 2.5\sigma$. A Gaussian fit is carried out in each of these peaks to provide an estimate of the source position and their flux densities are measured at the positions of those peaks in all the SPIRE bands. Finally, only sources with S/N $> 4\sigma$ in at least one of the three SPIRE bands are kept in the final catalog.

We have worked with the backsub maps instead of the fbacksub ones in order to test our own multifrequency matched-filter's performance without any other alteration but the subtraction of the large-scale background emission. The method used to subtract this large-scale emission does not affect the flux density of point sources. The units of the maps are Jy/beam. We converted these fluxes to Jy/pixel by dividing the values in the maps by the ratio between the beam area and the pixel area in arcsec^2 (469/36, 831/64 and 1804/144 at 250, 350 and 500 μm, respectively). The maps have pixel sizes of 6, 8 and 12 arcsec at 250, 350 and 500 μm, respectively. All maps must have the same pixel size so as to be able to combine the three-channel images of a source into one single filtered image. Thus we re-binned 350 and 500 μm maps to a pixel size of 6 arcsec, the same pixel size as the 250 μm map. This repixelization may cause small alignment errors between the pixel positions of the center of the source in the different channels. These pixel misalignments have already been considered and monitored in the method. We have achieved a perfect alignment for $\sim 90\%$ of the H-ATLAS sources, leaving the rest with deviations not greater than 2 pixels in one or some of the channels.

2.2 Choosing the frequency dependence

The frequency dependence f_ν of the sources is not known a priori just with the information of the images. For that reason, we have to model the spectral behavior of the sources somehow. The following template model from Pearson et al. (2013) has been used as frequency dependence for all the sources considered:

$$S_\nu = A_n \left[B_{\nu(1+z)}(T_h)[\nu \cdot (1+z)]^\beta + a B_{\nu(1+z)}(T_c)[\nu \cdot (1+z)]^\beta \right] \tag{2.1}$$

where S_ν is the flux at a redshift frequency $\nu(1+z)$, z is the unknown redshift of the source, A_n is a normalization factor, $B_{\nu(1+z)}$ is the Planck function, $\beta = 2$ is the emissivity index, $T_h = 46.9$ K and $T_c = 23.9$ K are the temperatures of the hot and cold dust components, and $a = 30.1$ is the ratio between cold and hot dust masses.

This template has been developed to estimate redshifts using only the SPIRE fluxes from *Herschel*. It has emerged from a subset of 40 bright *Herschel*-ATLAS sources with very well known redshifts in the range $0.5 < z < 4.3$. The redshifts of 25 of them, with $z < 1$, were obtained through optical spectroscopy. The redshifts of the other 15 objects, in the range $0.8 < z < 4.3$, were estimated from CO observations. This SED has already been used and studied in several previous works, among which we can cite Eales (2015); Ivison et al. (2016); Bianchini et al. (2016, 2018); Negrello et al. (2017); Fudamoto et al. (2017); Bakx et al. (2018); Donevski et al. (2018).

Given that all the sources used to build this model are among the most luminous H-ATLAS sources at their respective redshifts, a bias may arise from the fact that the model may not be representative of the less luminous sources. For instance, low-z H-ATLAS sources have cooler SEDs than the template derived in Pearson et al. (2013) from their high-z spectroscopic sample. It is important to bear in mind that the many different types of sources distributed in the sky constitute a very heterogeneous set of objects that do not have a common spectral behavior. This is the reason why the detection and estimation of the flux of point sources is a difficult task. In this sense, it should be noted that this template model is not expected to be a physically real SED but simply a representative model that can be used as a statistical tool for estimating redshifts from SPIRE fluxes.

2.3 Implementation of the MMF

Once all maps have the same pixel size and are in units of Jy/pixel, we can apply the MMF on the positions of the 4σ detections produced by the MADX algorithm. Positions (α and δ) of all point sources identified in the maps are known and taken from the H-ATLAS catalog, converted from degrees into pixels and used to select the objects in the backsub maps. We extract square patches of 256×256 pixels centered on the position of the source for the three bands (250, 350 and 500 μm) and align them to run our MMF algorithm. When a H-ATLAS source is close to the edge of the H-ATLAS footprint, the zeros in the map are replaced by white noise generated with the same mean and standard deviation as that of the rest of the map (within the footprint). The MMF takes as an argument this set of $N = 3$ images corresponding to the same area of the sky and returns a single filtered image where the source is optimally enhanced with respect to the noise. For N images, the frequency dependence f_ν has N degrees of freedom. Choosing one of the frequencies under consideration as fiducial frequency of reference allows reducing to N-1 the number of independent degrees of freedom. We decided to choose the 250 μm channel as fiducial frequency.

The final filtered image is the result of a two-phase process. The first phase is the slowest one but, having separated the dependence in q from the dependence in ν in Eq. (1.35), it only needs to be done once for each set of images of the source considered. It consists of the calculation of a prefiltered map without any frequency dependence information. This prefiltered map is built using the Fourier transforms of the $N = 3$ image patches and the matrix $\mathbf{H}$. The cross-power spectrum of the images and matrices $\mathbf{T}$ are needed in the calculation of the matrix $\mathbf{H}$. $\mathbf{T}$ matrices are derived using the corresponding PSFs at 250, 350 and 500 μm as source profiles at each frequency. And the inverse of the cross-power spectrum matrix is calculated for every pixel in the images.

The second phase is faster and requires only the calculation of the normalization factor α and the linear combination of the prefiltered maps using the frequency dependence f_ν shown in Eq.(2.1). For each source we seek to maximize the S/N of the final filtered map with respect to the frequency dependence f_ν. For the frequency dependence considered, the only free parameters in this optimization process are the source amplitude A and its redshift z. In fact, the amplitude is not really a variable, because for any given set of images it is determined by z for any iteration of the filter through Eq. (1.34). By construction, the resulting A coincides with the source's flux density when the optimization is completed. Therefore, the only variable in the optimization is z and the maximization of the filtered S/N of a given source is tantamount to finding its redshift. Since the redshift z of the

source is unknown, this last step is repeated for equally distributed redshift values in the range $0 \leq z \leq 7$, with a step of 0.007 until the redshift which maximizes the S/N defined in Eq. (1.8), that is, the optimal redshift, for the considered source is found. In the end, we have an estimation of the redshift of the source, its frequency dependence vector f_ν, and a maximized filtered image of the source, whose amplitude A corresponds directly to the source flux density in the chosen fiducial frequency (250 μm). The fluxes of the source at the other two frequencies (350 and 500 μm) can be obtained by multiplying this amplitude by the corresponding components of f_ν, which is normalized to the fiducial frequency. This method is robust only in the case of point sources, that is, those sources whose spatial profile in each frequency agrees with the beam profile in that frequency.

2.4 Simulations

In order to test our method, we used simulated data with a well known SED. Simulations are useful for both, identifying possible biases and estimating the errors in the reconstructed redshift. Simulations were done using only GAMA's backsub maps from H-ATLAS DR1. The recently released NGP and SGP fields from H-ATLAS DR2 were not used for the simulations, but this should have no impact on our results.

We started each simulation with a randomly chosen square patch of the desired size (256×256 pixels in our case) from any of the three equatorial fields surveyed (GAMA9, GAMA12 or GAMA15). The same patch region was selected for the three sub-mm photometric bands. Since the three SPIRE channels have different pixel sizes, and the MMF needs to work with a common pixel size, we re-bin the 350 and 500 μm maps to have the same pixel size as the 250 μm map. Alignment errors between the pixel positions of the source in the different channels (which may harm the MMF filtering result), can take place due to this repixelization but, as we already explained in Sec. 2.1, they have already been considered for the H-ATLAS sources, as they are for simulated sources. Thus, all maps used in simulations have a pixel size of 6 arcsec and are in units of Jy/pixel. Then a source with the corresponding beam profile (according to the PSF of the channel), an adequate amplitude (in order to obtain fluxes like the H-ATLAS ones), a fixed redshift, and known SED (Eq. 2.1), is placed in the middle of each one of the three patches. From this moment we followed the same procedure, described in Sec. 2.3, as with any H-ATLAS source. As it is done with the H-ATLAS sources, if the selected map patch contains zeros (i.e., it is near the edge of one of the GAMA fields), these are replaced by white noise with dispersion given by the map background. We performed 5000 simulations, as described before, for each one of the redshifts considered within the ranges $1 \leq z \leq 4.5$, with a 0.1 step, and

$4.5 \leq z \leq 7$, with a 0.5 step. For each input redshift value considered, z_{in}, we compute the mean value of the 5000 output redshift values, z_{out}, and the standard deviation. The difference between the redshifts estimated after applying our MMF method to the simulated sources and the input redshift is shown in Fig. 2.1 as a function of z_{in}.

The bias observed above $z \gtrapprox 4$ could be due (in part) to the fact that the Pearson model (Eq. 2.1) is built based on some of the most luminous H-ATLAS sources and a restricted range in redshift ($0.5 \leq z \leq 4.3$). However, more importantly, photometric redshifts derived from a SED have problems when the peak of the IR emission is not bracketed by the three SPIRE bands. This peak falls in the SPIRE bands between $z \approx 1$ and $z \approx 4$, and it is precisely in this redshift range where our method seems more robust returning unbiased redshift estimates. Beyond $z \approx 4$ a positive bias can be appreciated which can be as high as ≈ 0.6 at $z \approx 7$. Using a polynomial fit, we find that our estimations of the redshift after applying the MMF can be corrected through:

$$z_{true} = 0.0013 \cdot z_{out}^4 - 0.018 \cdot z_{out}^3 + 0.056 \cdot z_{out}^2 + 0.943 \cdot z_{out} + 0.057 \tag{2.2}$$

where z_{true} is the unbiased redshift estimation of the corresponding H-ATLAS source.

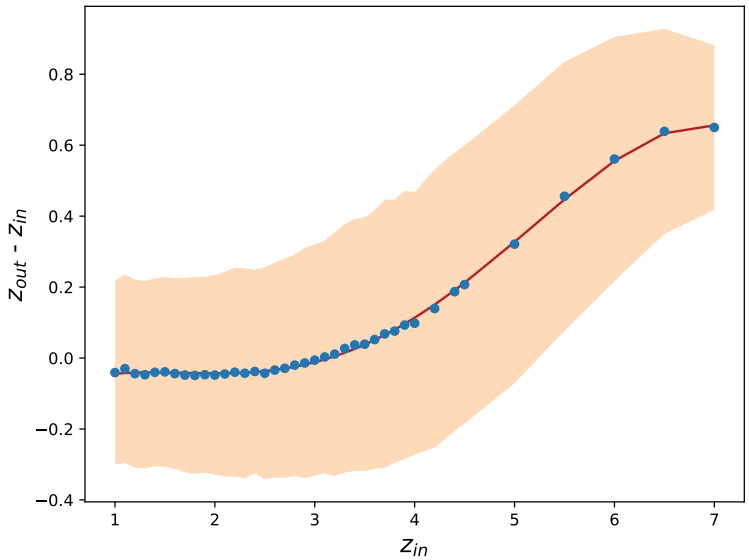


Figure 2.1: Difference between the redshift recovered with our MMF method (z_{out}) and the input redshift (z_{in}) as a function of z_{in}. 5000 simulations are run for each z_{in} value in the range $1 \leq z_{in} \leq 7$. The mean value (z_{out}) is computed and shown as blue dots. The solid red line shows a polynomial fit to these mean values. This fit is later used to correct for this bias. z_{out}-z_{in} errors (1σ) are plotted as a shaded region.

2.5 Comparison with known-redshift H-ATLAS sources

We compare the redshifts obtained by the MMF method with a set of 32 *Herschel*-ATLAS sources with known spectroscopic redshifts from Pearson et al. (2013), Negrello et al. (2017) and Bakx et al. (2018). Several of the sources selected are ubiquitous in all these references. Ten of these 32 sources are chosen from Negrello et al. (2017), 17 are sources with $z_{spec} > 0.8$ used in Pearson et al. (2013) to build their template and five are taken from Bakx et al. (2018). Redshifts and flux densities estimated with the MMF for these sources are shown in Table 2.1.

The differences between photometric redshifts estimated with the MMF and the measured spectroscopic redshifts for these 32 objects are shown in Fig. 2.2. The top plot shows $\Delta z/(1+z_{spec}) = (z_{photo} - z_{spec})/(1+z_{spec})$ before the bias correction. The mean and median are $\mu = 0.004$ and $\mu_{1/2}$ = -0.017, respectively, with a *rms* scatter of $\sigma = 0.143$. The bottom plot shows the same quantity after the bias correction. Since at $z < 4$ the bias correction is small, the improvement is small in this redshift range. Nevertheless, the mean and median ($\mu = 0.009$, $\mu_{1/2}$ = -0.003), and scatter ($\sigma = 0.138$) are slightly better than in the sample without bias correction. These statistical parameters are also included in the corresponding redshift column of Table 2.1. If we take the definition for outliers (those with $|\Delta z / (1+z_{spec})| > 0.3$) used in Ivison et al. (2016), only one of the objects considered is identified as an outlier both for the analysis with biased and unbiased redshifts. Error bars in the top panel are calculated by using Eqs. 1.34 and 2.1, while error bars in the bottom panel are derived from simulations described in Sec. 2.4. We note how the error bars in the bottom panel are more representative of the dispersion around the zero value than the error bars in the top plot. This result indirectly confirms that the error bars derived from the simulations are the most meaningful ones for our estimated redshifts.

Despite this result, it is worth to clarify something more about the errors here calculated. From Eq. (1.34), and using a parametric SED such as Eq. (2.1) it is possible, under some general (but not necessarily true) assumptions, such as the statistical independence of the background noise at the three different SPIRE channels, to estimate the degree of uncertainty of our photometric redshift z_{MMF}. The error bars of all redshift estimations from tables 2.1 to 2.3, except for $z_{MMF}^{unbiased}$, have been obtained this way. However, the statistical *uncertainty* of an estimator and its actual *error* with respect to the groundtruth are not necessarily the same thing. The uncertainty given to an estimator can be under, or overestimated depending on the validity of the statistical assumptions made. On the other hand, the estimator may be biased and this bias may not be accounted for in the calculation of the uncertainty. When possible, it is preferable to calculate the error of the

estimator using real, already known values of z_{spec} or, if few spectroscopic redshifts are available, by means of realistic simulations. This is the approach followed in this section to obtain the errors of the unbiased redshifts $z_{MMF}^{unbiased}$ from Table 2.1, showing that for SPIRE, and in the redshift range $1 \leq z \leq 7$, the actual average error of the estimation of z is under control and typically smaller than the uncertainty calculated from Eq. (1.34) and Eq. (2.1).

In order to test the robustness of our results, we have repeated the comparison with the spectroscopic sources, but changing the maps from which we extract these sources and the convolution functions used. We run our method, but using several combinations of the backsub and fbacksub maps as well as SPIRE's PSFs and MFs. The results obtained with the different configurations are shown in Table 2.2. We also obtain redshift estimates for these configurations by applying the Pearson's χ^2 test statistic, but without using the MMF, and taking into account only the flux density measurements from different maps or the tabulated fluxes from the H-ATLAS catalog, and comparing them with the fluxes predicted by the Pearson et al. (2013) SED. These last results are shown in Table 2.3.

As can be seen by comparing the redshift estimations for the 32 known-spectroscopic sources from tables 2.1 to 2.3, both our unbiased and biased estimates, obtained with the MMF method on the backsub maps and using the SPIRE PSFs, outperform all the other redshift estimations derived with the alternative combinations of maps and convolution functions. Only redshifts estimated by applying a Pearson's χ^2 test statistic with flux measurements from fbacksub maps ($z_{\chi^2}^{fb}$) and with tabulated fluxes from the H-ATLAS catalog ($z_{\chi^2}^{t}$) get comparable results. Nevertheless, our unbiased MMF redshifts are the ones which get lower offsets and scatter, and agree with spectroscopic redshifts with the greatest accuracy.

Focusing now our attention on flux densities, practically all fluxes recovered with the MMF (S_{MMF}) are consistent with the corresponding tabulated fluxes from H-ATLAS catalog (S_{SPIRE}), in the sense that the IR peak is recovered at the right corresponding wavelength for 29 out of the 32 sources considered from Table 2.1. On the other hand, and as expected, all IR peaks are recovered in the right band for the flux densities $S_{\chi^2}^{fb}$ taken from the fbacksub maps (see Table 2.3). The comparison between our MMF estimates of the flux densities and those from the H-ATLAS catalog is shown in Fig. 2.3. It can be seen how our flux estimations seem to be systematically below the values from the H-ATLAS catalog. This slight underestimate is expected since the noise reduction carried out by the MMF must lead to flux densities lower than those of H-ATLAS. The average flux underestimates between the flux densities estimated with our MMF and the H-ATLAS fluxes are 17 ± 13 mJy (250 μm), 18 ± 9 mJy (350 μm) and 14 ± 14 mJy (500 μm).

Table 2.1: 32 sources used to test the matched multifilter (MMF) method which includes a combination of sources from Pearson et al. (2013), Negrello et al. (2017) and Bakx et al. (2018). For each source the following information is provided: measured spectroscopic redshift, *Herschel*/SPIRE fluxes as tabulated in the H-ATLAS catalog (S_{SPIRE}), our redshift estimate before (z_{MMF}) and after ($z_{MMF}^{unbiased}$) bias correction, and the MMF flux densities. The statistical parameters μ, $\mu_{1/2}$ and σ at the bottom of the table are the mean, median and rms scatter of the quantity $\Delta z/(1 + z_{spec})$ for the corresponding photometric redshift.

H-ATLAS name	α (deg)	δ (deg)	z_{spec}	S_{SPIRE} (mJy)			z_{MMF}	$z_{MMF}^{unbiased}$	S_{MMF} (mJy)		
				250 μm	350 μm	500 μm			250 μm	350 μm	500 μm
GAMA-9											
H-ATLAS J083051.0+013225	127.71268	1.54028	3.634◇	249 ± 7	305 ± 8	269 ± 9	3.15 ± 1.05	3.15 ± 0.34	230 ± 5	288 ± 7	265 ± 6
H-ATLAS J085358.9+015537	133.49574	1.92712	2.091⋆	396 ± 8	368 ± 8	228 ± 9	2.05 ± 0.22	2.09 ± 0.28	365 ± 7	361 ± 7	243 ± 5
H-ATLAS J091043.1-000321	137.67949	-0.0562	1.784⋆	421 ± 7	371 ± 7	221 ± 8	1.79 ± 0.17	1.83 ± 0.28	388 ± 7	354 ± 6	218 ± 4
H-ATLAS J090302.9-014127	135.76235	-1.69088	2.308⋆	354 ± 7	339 ± 8	220 ± 9	2.10 ± 0.24	2.14 ± 0.29	313 ± 6	314 ± 7	215 ± 4
H-ATLAS J090740.0-004200	136.91691	-0.70018	1.577⋆	478 ± 7	328 ± 8	171 ± 9	1.20 ± 0.07	1.24 ± 0.26	450 ± 8	325 ± 6	160 ± 3
H-ATLAS J090311.6+003906	135.79868	0.65206	3.037⋆	133 ± 7	186 ± 8	165 ± 9	3.17 ± 0.34	3.17 ± 0.34	139 ± 5	174 ± 6	162 ± 6
H-ATLAS J091305.0-005343	138.27077	-0.89563	2.626⋆	118 ± 6	137 ± 7	104 ± 8	2.50 ± 0.47	2.53 ± 0.30	123 ± 5	136 ± 6	105 ± 4
H-ATLAS J084933.4+021443	132.38934	2.24526	2.410⋆	217 ± 8	249 ± 8	209 ± 9	2.35 ± 0.47	2.39 ± 0.30	201 ± 6	216 ± 7	160 ± 5
H-ATLAS J091840.8+023047	139.67015	2.51348	2.581⋆	126 ± 7	151 ± 8	128 ± 9	2.94 ± 2.14	2.95 ± 0.33	109 ± 6	132 ± 7	116 ± 6
GAMA-12											
H-ATLAS J114637.9-001132	176.65816	-0.19226	3.259⋆	316 ± 7	358 ± 7	292 ± 8	2.53 ± 0.36	2.56 ± 0.31	289 ± 6	322 ± 6	252 ± 5
H-ATLAS J114023.0-001043	175.09587	-0.17810	0.844⋆	68 ± 7	40 ± 8	20 ± 9	0.78 ± 0.19	0.82 ± 0.27	64 ± 8	38 ± 5	16 ± 2
H-ATLAS J115820.2-013753	179.58398	-1.6313	2.191⋆	120 ± 7	124 ± 8	101 ± 8	2.40 ± 0.93	2.44 ± 0.30	94 ± 7	102 ± 7	77 ± 5
H-ATLAS J113243.0-005108	173.17953	-0.85248	2.578⋆	68 ± 7	106 ± 8	100 ± 9	3.69 ± 2.90	3.64 ± 0.36	59 ± 4	81 ± 6	83 ± 6
H-ATLAS J113526.3-014605	173.85958	-1.76854	3.128⋆	279 ± 7	283 ± 8	204 ± 9	2.21 ± 0.51	2.25 ± 0.29	260 ± 6	269 ± 6	191 ± 5
GAMA-15											
H-ATLAS J142935.3-002836	217.39717	-0.47682	1.026⋆	801 ± 7	438 ± 7	200 ± 8	0.59 ± 0.01	0.63 ± 0.28	769 ± 7	413 ± 4	161 ± 2
H-ATLAS J142413.9+022303	216.0582	2.38424	4.243⋆	112 ± 7	182 ± 8	193 ± 8	4.04 ± 0.76	3.94 ± 0.38	117 ± 4	169 ± 6	186 ± 7
H-ATLAS J141351.9-000026	213.46663	-0.00745	2.478⋆	189 ± 7	217 ± 8	176 ± 9	2.70 ± 0.65	2.73 ± 0.31	166 ± 5	192 ± 6	158 ± 5
H-ATLAS J141148.9-011439	212.95392	-1.24429	0.857⋆	61 ± 7	44 ± 8	17 ± 9	1.11 ± 0.57	1.15 ± 0.26	53 ± 12	37 ± 9	17 ± 4
NGP											
H-ATLAS J134429.4+303034	206.12285	30.50948	2.301◇	462 ± 8	466 ± 9	343 ± 9	2.25 ± 0.18	2.29 ± 0.29	443 ± 6	463 ± 7	332 ± 5
H-ATLAS J132630.1+334407	201.6255	33.73548	2.951◇	191 ± 7	281 ± 8	279 ± 9	4.04 ± 0.52	3.94 ± 0.38	171 ± 4	247 ± 6	270 ± 7
H-ATLAS J125632.4+233627	194.13524	23.60758	3.565◇	209 ± 7	288 ± 8	264 ± 9	3.41 ± 8	3.39 ± 0.35	201 ± 5	261 ± 7	256 ± 6
H-ATLAS J132427.0+284449	201.11257	28.74716	1.676◇	342 ± 7	371 ± 8	251 ± 9	2.23 ± 0.22	2.27 ± 0.29	329 ± 7	342 ± 7	244 ± 5
H-ATLAS J125135.3+261457	192.89721	26.24942	3.675◇	158 ± 8	202 ± 8	207 ± 9	4.47 ± 0.44	4.30 ± 0.39	112 ± 4	177 ± 6	206 ± 6
H-ATLAS J133008.5+245900	202.53579	24.98334	3.111◇	271 ± 7	278 ± 8	203 ± 8	2.29 ± 0.31	2.33 ± 0.29	255 ± 6	269 ± 7	196 ± 5
H-ATLAS J133649.9+291800	204.20819	29.3	2.202◇	294 ± 7	286 ± 8	194 ± 8	1.79 ± 0.17	1.83 ± 0.27	291 ± 7	265 ± 6	162 ± 4
H-ATLAS J133542.9+300401	203.92905	30.06709	2.685◇	137 ± 7	146 ± 8	125 ± 8	2.21 ± 0.60	2.25 ± 0.29	130 ± 6	134 ± 7	95 ± 5
H-ATLAS J132859.2+292327	202.2468	29.39074	2.778◇	268 ± 7	296 ± 7	249 ± 8	2.83 ± 0.44	2.85 ± 0.32	233 ± 5	275 ± 6	234 ± 5
H-ATLAS J125652.5+275900	194.21857	27.98339	2.79◇	134 ± 8	164 ± 8	132 ± 9	2.76 ± 0.66	2.78 ± 0.32	129 ± 5	151 ± 6	126 ± 5
H-ATLAS J130118.0+253708	195.32517	25.61902	4.04◇	60 ± 7	101 ± 8	102 ± 8	3.87 ± 2.23	3.79 ± 0.37	57 ± 4	80 ± 5	85 ± 6
H-ATLAS J132301.7+341649	200.75701	34.28036	2.19◇	124 ± 7	145 ± 8	137 ± 9	3.35 ± 9	3.33 ± 0.35	102 ± 5	131 ± 6	127 ± 6
H-ATLAS J132504.4+311534	201.26821	31.25951	1.84◇	241 ± 7	227 ± 8	165 ± 9	1.79 ± 0.21	1.83 ± 0.27	233 ± 8	212 ± 7	130 ± 4
H-ATLAS J133846.5+255055	204.69389	25.84855	2.34◇	159 ± 8	183 ± 8	138 ± 9	2.42 ± 0.61	2.46 ± 0.30	151 ± 7	165 ± 7	125 ± 5
$\mu_{1/2}$							-0.017	-0.003			
μ							0.004	0.009			
σ							0.143	0.138			

⋆ from Pearson et al. (2013)
◇ from Negrello et al. (2017)
○ from Bakx et al. (2018)

Table 2.2: As in Table 2.1 but for fluxes and redshifts estimated using the backsub maps and convolving the images of the sources with the SPIRE MFs (z^{b}_{MF} and S^{b}_{MF}), or using the fbacksub maps and convolving the images of the sources with the SPIRE MFs (z^{fb}_{MF} and S^{fb}_{MF}), or using the fbacksub maps and convolving the images of the sources with the SPIRE PSFs (z^{fb}_{PSF} and S^{fb}_{PSF}).

H-ATLAS name	z_{spec}	z^{b}_{MF}	S^{b}_{MF} (mJy)			z^{fb}_{MF}	S^{fb}_{MF} (mJy)			z^{fb}_{PSF}	S^{fb}_{PSF} (mJy)		
			250 μm	350 μm	500 μm		250 μm	350 μm	500 μm		250 μm	350 μm	500 μm
GAMA-9													
H-ATLAS J083051.0+013225	3.634◇	2.50 ± 0.54	186 ± 5	206 ± 5	160 ± 4	2.13 ± 0.28	117 ± 4	119 ± 4	82 ± 3	3.19 ± 1.45	181 ± 5	228 ± 6	212 ± 5
H-ATLAS J085358.9+015537	2.091⋆	1.47 ± 0.09	288 ± 6	233 ± 5	127 ± 3	1.14 ± 0.05	175 ± 4	123 ± 3	59 ± 2	1.89 ± 0.22	288 ± 5	272 ± 5	173 ± 3
H-ATLAS J091043.1-000321	1.784⋆	1.20 ± 0.05	290 ± 6	209 ± 4	103 ± 2	0.95 ± 0.05	113 ± 3	72 ± 2	32 ± 1	1.58 ± 0.15	216 ± 5	182 ± 4	104 ± 2
H-ATLAS J090302.9-014127	2.308⋆	1.64 ± 0.12	236 ± 6	204 ± 5	118 ± 3	1.23 ± 0.09	116 ± 4	85 ± 3	42 ± 2	1.97 ± 0.27	215 ± 5	208 ± 5	136 ± 3
H-ATLAS J090740.0-004200	1.577⋆	0.62 ± 0.02	347 ± 7	189 ± 4	75 ± 2	0 ± 0.01	202 ± 4	80 ± 2	25 ± 1	0.87 ± 0.03	349 ± 6	215 ± 4	93 ± 2
H-ATLAS J090311.6+003906	3.037⋆	2.96 ± 26	103 ± 4	125 ± 5	110 ± 4	3.47 ± 0.57	48 ± 2	64 ± 3	63 ± 3	3.57 ± 0.60	96 ± 4	129 ± 5	130 ± 5
H-ATLAS J091305.0-005343	2.626⋆	2.02 ± 0.33	91 ± 4	89 ± 4	59 ± 3	3.24 ± 1.01	30 ± 2	38 ± 2	35 ± 2	3.38 ± 1.24	62 ± 3	80 ± 4	78 ± 4
H-ATLAS J084933.4+021443	2.410⋆	1.98 ± 0.31	138 ± 5	134 ± 5	88 ± 3	1.91 ± 0.42	71 ± 3	67 ± 3	43 ± 2	2.49 ± 1.38	141 ± 5	156 ± 5	120 ± 4
H-ATLAS J091840.8+023047	2.581⋆	2.80 ± 5	76 ± 5	90 ± 6	76 ± 5	3.14 ± 1.32	35 ± 3	44 ± 3	40 ± 3	3.31 ± 1.35	73 ± 4	93 ± 6	89 ± 5
GAMA-12													
H-ATLAS J114637.9-001132	3.259⋆	1.97 ± 0.17	201 ± 5	194 ± 4	127 ± 3	2.24 ± 0.81	81 ± 2	84 ± 2	60 ±	2.82 ± 1.11	155 ± 4	184 ± 4	156 ± 4
H-ATLAS J114023.0-001043	0.844⋆	0.27 ± 0.08	51 ± 7	23 ± 3	8 ± 1	0.58 ± 0.19	22 ± 4	12 ± 2	5 ± 1	0.69 ± 0.18	43 ± 6	24 ± 4	10 ± 2
H-ATLAS J115820.2-013753	2.191⋆	2.27 ± 0.91	61 ± 5	64 ± 5	46 ± 4	2.31 ± 2	33 ± 3	35 ± 3	26 ± 2	2.61 ± 4.5	66 ± 4	75 ± 5	60 ± 4
H-ATLAS J113243.0-005108	2.578⋆	3.06 ± 14	42 ± 4	52 ± 4	47 ± 4	2.77 ± 2.18	30 ± 3	35 ± 4	29 ± 3	3.85 ± 3	51 ± 4	72 ± 5	76 ± 5
H-ATLAS J113526.3-014605	3.128⋆	1.79 ± 0.19	198 ± 5	181 ± 5	111 ± 3	1.78 ± 0.24	106 ± 3	97 ± 3	59 ± 2	2.07 ± 0.56	190 ± 5	189 ± 5	128 ± 3
GAMA-15													
H-ATLAS J142935.3-002836	1.026⋆	0 ± 0.01	578 ± 6	230 ± 2	72 ± 1	0 ± 0.01	247 ± 3	98 ± 1	31 ± 1	0.18 ± 0.01	469 ± 5	204 ± 2	68 ± 1
H-ATLAS J142413.9+022303	4.243⋆	3.89 ± 0.57	86 ± 4	122 ± 5	130 ± 5	4.72 ± 0.38	34 ± 2	56 ± 3	67 ± 4	4.83 ± 0.36	69 ± 3	117 ± 5	143 ± 6
H-ATLAS J141351.9-000026	2.478⋆	2.36 ± 0.53	119 ± 5	128 ± 5	95 ± 4	2.12 ± 0.76	60 ± 3	62 ± 3	42 ± 2	2.89 ± 5.6	112 ± 4	134 ± 5	115 ± 4
H-ATLAS J141148.9-011439	0.857⋆	1.30 ± 0.59	35 ± 8	27 ± 6	14 ± 3	1.12 ± 0.50	20 ± 4	14 ± 3	7 ± 1	1.42 ± 0.75	37 ± 8	29 ± 6	16 ± 3
NGP													
H-ATLAS J134429.4+303034	2.301◇	1.72 ± 0.09	341 ± 6	304 ± 5	182 ± 3	2.03 ± 0.27	171 ± 3	169 ± 3	112 ± 2	2.40 ± 0.79	312 ± 5	338 ± 5	254 ± 4
H-ATLAS J132630.1+334407	2.951◇	3.86 ± 0.45	118 ± 3	165 ± 5	176 ± 5	5.73 ± 0.25	35 ± 1	75 ± 3	103 ± 4	5.40 ± 0.23	84 ± 2	165 ± 5	216 ± 6
H-ATLAS J125632.4+233627	3.565◇	3.27 ± 1.46	144 ± 4	184 ± 5	174 ± 5	4.97 ± 0.26	47 ± 2	82 ± 3	102 ± 4	4.93 ± 0.26	97 ± 3	169 ± 5	209 ± 6
H-ATLAS J132427.0+284449	1.676◇	1.87 ± 0.16	241 ± 6	226 ± 5	142 ± 3	2.00 ± 0.38	103 ± 3	100 ± 3	66 ± 2	2.31 ± 0.53	205 ± 5	217 ± 5	159 ± 4
H-ATLAS J125135.3+261457	3.675◇	4.16 ± 0.43	84 ± 3	124 ± 5	139 ± 5	4.27 ± 0.37	44 ± 2	67 ± 3	76 ± 3	4.67 ± 0.33	84 ± 3	137 ± 5	164 ± 6
H-ATLAS J133008.5+245900	3.111◇	1.84 ± 0.18	195 ± 5	181 ± 5	113 ± 3	2.33 ± 1.52	69 ± 3	74 ± 3	54 ± 2	2.46 ± 0.60	131 ± 4	144 ± 5	111 ± 4
H-ATLAS J133649.9+291800	2.202◇	1.44 ± 0.10	212 ± 6	169 ± 5	91 ± 3	1.84 ± 0.32	75 ± 3	70 ± 3	44 ± 2	2.03 ± 0.41	160 ± 5	157 ± 5	105 ± 3
H-ATLAS J133542.9+300401	2.685◇	2.21 ± 0.64	89 ± 5	92 ± 5	65 ± 4	3.13 ± 1.53	32 ± 3	40 ± 3	36 ± 3	2.68 ± 2.46	72 ± 4	83 ± 5	68 ± 4
H-ATLAS J132859.2+292327	2.778◇	2.21 ± 0.23	177 ± 4	183 ± 4	129 ± 3	2.19 ± 0.62	78 ± 2	81 ± 2	57 ± 2	2.96 ± 1.15	146 ± 4	177 ± 4	156 ± 4
H-ATLAS J125652.5+275900	2.79◇	2.45 ± 0.66	98 ± 5	108 ± 5	82 ± 4	3.34 ± 0.89	46 ± 3	60 ± 3	57 ± 3	3.26 ± 1.75	91 ± 4	116 ± 5	110 ± 5
H-ATLAS J130118.0+253708	4.04◇	6.24 ± 0.60	17 ± 1	43 ± 3	62 ± 5	5.62 ± 0.54	15 ± 1	30 ± 2	41 ± 3	5.23 ± 0.60	34 ± 2	63 ± 4	81 ± 5
H-ATLAS J132301.7+341649	2.19◇	2.90 ± 5	81 ± 4	97 ± 5	84 ± 4	3.28 ± 1.23	32 ± 2	41 ± 3	39 ± 3	4.29 ± 0.74	55 ± 3	83 ± 5	94 ± 6
H-ATLAS J132504.4+311534	1.84◇	1.46 ± 0.14	170 ± 6	137 ± 5	75 ± 3	2.33 ± 2.12	62 ± 3	66 ± 3	49 ± 3	2.42 ± 0.66	118 ± 5	128 ± 5	97 ± 4
H-ATLAS J133846.5+255055	2.34◇	2.07 ± 0.42	115 ± 5	115 ± 5	78 ± 4	3.13 ± 1.24	46 ± 3	57 ± 4	52 ± 3	3.18 ± 1.88	94 ± 4	119 ± 6	110 ± 5
$\mu_{1/2}$		-0.128				0.045				0.115			
μ		-0.092				-0.016				0.088			
σ		0.201				0.272				0.230			

⋆ from Pearson et al. (2013)
◇ from Negrello et al. (2017)
◦ from Bakx et al. (2018)

Table 2.3: As in Table 2.1 but for redshifts estimated by applying a χ^2 test statistic considering flux measurements from backsub ($S^b_{\chi^2}$) and fbacksub ($S^{fb}_{\chi^2}$) maps, respectively, as the observed data and the fluxes obtained directly with the SED from Eq. (2.1) as the theoretical data ($z^b_{\chi^2}$ and $z^{fb}_{\chi^2}$); and by applying a χ^2 statistic considering the tabulated fluxes from the H-ATLAS catalog as the observed data and the fluxes obtained directly with the same SED as the theoretical data ($z^t_{\chi^2}$).

H-ATLAS name	z_{spec}	$z^b_{\chi^2}$	$S^b_{\chi^2}$ (mJy)			$z^{fb}_{\chi^2}$	$S^{fb}_{\chi^2}$ (mJy)			$z^t_{\chi^2}$
			250 μm	350 μm	500 μm		250 μm	350 μm	500 μm	
GAMA-9										
H-ATLAS J083051.0+013225	3.634°	3.22 ± 1	219 ± 12	283 ± 11	258 ± 12	3.02 ± 1	247 ± 9	303 ± 9	269 ± 9	3.01 ± 1
H-ATLAS J085358.9+015537	2.091⋆	2.02 ± 1	348 ± 12	347 ± 11	223 ± 12	1.82 ± 1	394 ± 9	356 ± 8	227 ± 9	1.84 ± 1
H-ATLAS J091043.1-000321	1.784⋆	1.66 ± 1	406 ± 10	339 ± 9	214 ± 10	1.70 ± 1	416 ± 8	361 ± 8	220 ± 8	1.70 ± 1
H-ATLAS J090302.9-014127	2.308⋆	1.99 ± 1	316 ± 11	296 ± 10	207 ± 11	2.00 ± 1	337 ± 8	327 ± 8	218 ± 8	1.94 ± 1
H-ATLAS J090740.0-004200	1.577⋆	1.35 ± 1	424 ± 12	321 ± 11	172 ± 11	1.18 ± 1	467 ± 9	323 ± 9	169 ± 9	1.16 ± 1
H-ATLAS J090311.6+003906	3.037⋆	3.81 ± 1	114 ± 11	171 ± 10	161 ± 11	3.48 ± 1	131 ± 8	184 ± 8	166 ± 8	3.43 ± 1
H-ATLAS J091305.0-005343	2.626⋆	2.47 ± 1	119 ± 9	132 ± 8	100 ± 9	2.59 ± 1	117 ± 7	135 ± 7	105 ± 7	2.58 ± 1
H-ATLAS J084933.4+021443	2.410⋆	2.95 ± 1	183 ± 11	223 ± 10	193 ± 11	2.71 ± 1	214 ± 8	246 ± 8	206 ± 8	2.72 ± 1
H-ATLAS J091840.8+023047	2.581⋆	3.19 ± 1	118 ± 12	150 ± 12	138 ± 12	2.99 ± 1	120 ± 10	151 ± 10	128 ± 10	2.87 ± 1
GAMA-12										
H-ATLAS J114637.9-001132	3.259⋆	2.64 ± 1	293 ± 9	333 ± 9	271 ± 10	2.63 ± 1	313 ± 7	346 ± 8	291 ± 9	2.64 ± 1
H-ATLAS J114023.0-001043	0.844⋆	1.24 ± 1	71 ± 11	46 ± 10	29 ± 11	0.95 ± 1	66 ± 8	44 ± 8	20 ± 8	0.91 ± 1
H-ATLAS J115820.2-013753	2.191⋆	2.79 ± 1	101 ± 11	115 ± 10	101 ± 11	2.52 ± 1	115 ± 10	124 ± 9	101 ± 10	2.42 ± 1
H-ATLAS J113243.0-005108	2.578⋆	3.34 ± 1	77 ± 12	100 ± 9	96 ± 10	4.03 ± 1	67 ± 9	106 ± 8	101 ± 8	3.97 ± 1
H-ATLAS J113526.3-014605	3.128⋆	2.40 ± 1	240 ± 11	263 ± 10	195 ± 12	2.31 ± 1	265 ± 7	287 ± 8	204 ± 9	2.19 ± 1
GAMA-15										
H-ATLAS J142935.3-002836	1.026⋆	0.79 ± 1	733 ± 9	415 ± 9	195 ± 10	0.77 ± 1	782 ± 8	428 ± 8	206 ± 8	0.72 ± 1
H-ATLAS J142413.9+022303	4.243⋆	4.18 ± 1	111 ± 11	162 ± 11	185 ± 12	4.35 ± 1	111 ± 8	177 ± 9	192 ± 9	4.37 ± 1
H-ATLAS J141351.9-000026	2.478⋆	2.84 ± 1	164 ± 11	196 ± 10	165 ± 11	2.78 ± 1	179 ± 8	212 ± 8	175 ± 9	2.67 ± 1
H-ATLAS J141148.9-011439	0.857⋆	1.81 ± 1	57 ± 20	48 ± 19	35 ± 18	1.11 ± 1	59 ± 19	44 ± 18	18 ± 17	1.06 ± 1
NGP										
H-ATLAS J134429.4+303034	2.301°	2.06 ± 1	459 ± 12	439 ± 11	313 ± 11	2.22 ± 1	457 ± 8	463 ± 8	340 ± 8	2.21 ± 1
H-ATLAS J132630.1+334407	2.951°	3.59 ± 1	187 ± 12	246 ± 11	256 ± 11	3.95 ± 1	186 ± 9	279 ± 8	278 ± 8	3.87 ± 1
H-ATLAS J125632.4+233627	3.565°	3.15 ± 1	216 ± 12	267 ± 11	251 ± 11	3.54 ± 1	202 ± 9	287 ± 9	261 ± 9	3.45 ± 1
H-ATLAS J132427.0+284449	1.676°	2.13 ± 1	338 ± 12	344 ± 11	235 ± 12	2.23 ± 1	339 ± 10	357 ± 9	250 ± 9	2.25 ± 1
H-ATLAS J125135.3+261457	3.675°	3.39 ± 1	148 ± 12	177 ± 11	193 ± 11	3.61 ± 1	150 ± 9	201 ± 8	206 ± 9	3.46 ± 1
H-ATLAS J133008.5+245900	3.111°	2.21 ± 1	259 ± 12	257 ± 11	194 ± 11	2.29 ± 1	264 ± 8	274 ± 8	204 ± 8	2.24 ± 1
H-ATLAS J133649.9+291800	2.202°	1.97 ± 1	277 ± 10	267 ± 10	174 ± 11	2.02 ± 1	290 ± 8	281 ± 9	192 ± 9	2.03 ± 1
H-ATLAS J133542.9+300401	2.685°	3.04 ± 1	117 ± 11	149 ± 10	127 ± 11	2.60 ± 1	135 ± 8	143 ± 8	125 ± 8	2.57 ± 1
H-ATLAS J132859.2+292327	2.778°	2.80 ± 1	239 ± 9	281 ± 9	237 ± 9	2.68 ± 1	258 ± 8	286 ± 8	247 ± 7	2.61 ± 1
H-ATLAS J125652.5+275900	2.79°	3.04 ± 1	122 ± 11	143 ± 11	138 ± 11	2.98 ± 1	128 ± 8	162 ± 8	136 ± 8	2.82 ± 1
H-ATLAS J130118.0+253708	4.04°	4.34 ± 1	56 ± 10	99 ± 9	89 ± 10	4.39 ± 1	59 ± 8	100 ± 8	101 ± 8	4.37 ± 1
H-ATLAS J132301.7+341649	2.19°	2.77 ± 1	126 ± 12	131 ± 11	130 ± 11	2.99 ± 1	123 ± 9	140 ± 9	136 ± 9	2.99 ± 1
H-ATLAS J132504.4+311534	1.84°	2.21 ± 1	206 ± 12	207 ± 11	152 ± 12	2.05 ± 1	237 ± 8	218 ± 9	164 ± 9	2.05 ± 1
H-ATLAS J133846.5+255055	2.34°	2.51 ± 1	146 ± 12	162 ± 11	127 ± 12	2.59 ± 1	156 ± 9	184 ± 9	136 ± 9	2.54 ± 1
$\mu_{1/2}$		0.028				0.008				-0.002
μ		0.041				0.024				0.011
σ		0.151				0.137				0.135

⋆ from Pearson et al. (2013)
◇ from Negrello et al. (2017)
○ from Bakx et al. (2018)

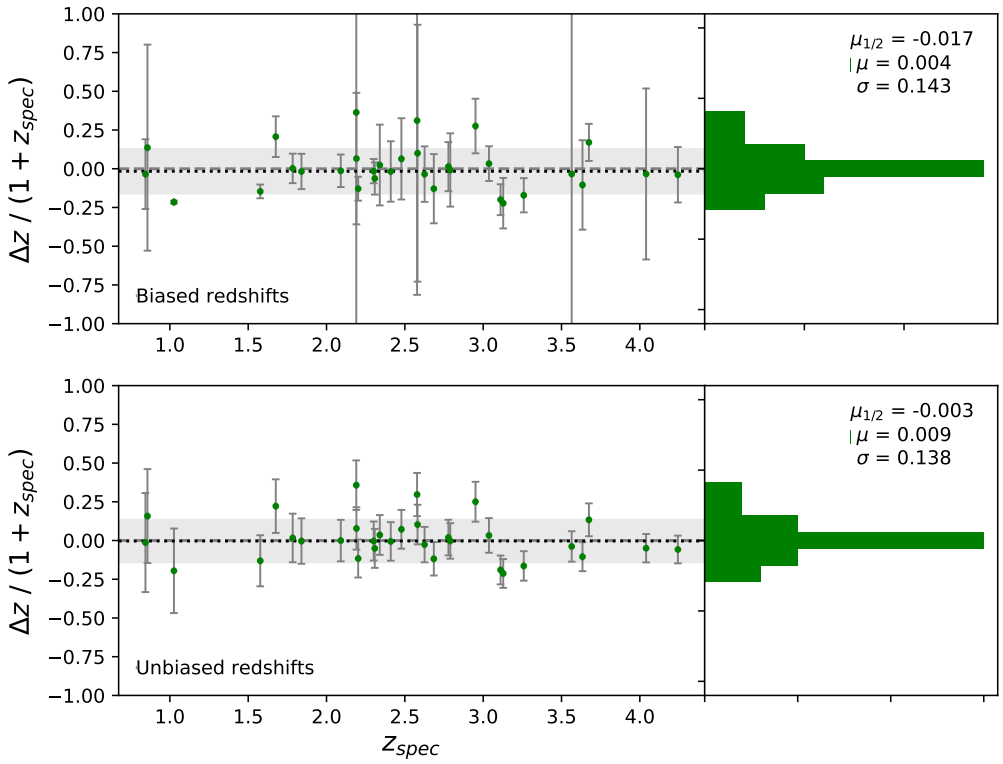


Figure 2.2: Difference, $\Delta z/(1 + z_{spec})$, as a function of z_{spec} between the biased (top) or unbiased (bottom) photometric redshifts estimated with our matched multifilter (MMF) and the spectroscopic redshifts from sources in Table 2.1. The statistical parameters noted illustrate the systematic overestimates or underestimates, mean μ and median $\mu_{1/2}$, and the degree of scatter, σ, of the photometric redshifts (z_{MMF}^{biased} and $z_{MMF}^{unbiased}$) found using the MMF and the Pearson et al. (2013) SED. Error bars in the top panel are calculated from using Eq. (1.34) and Eq. (2.1) while error bars in the bottom panel are derived from simulations described in Sec. 2.4. The gray band depicts the 1σ region centered in the median $\mu_{1/2}$ value for each distribution.

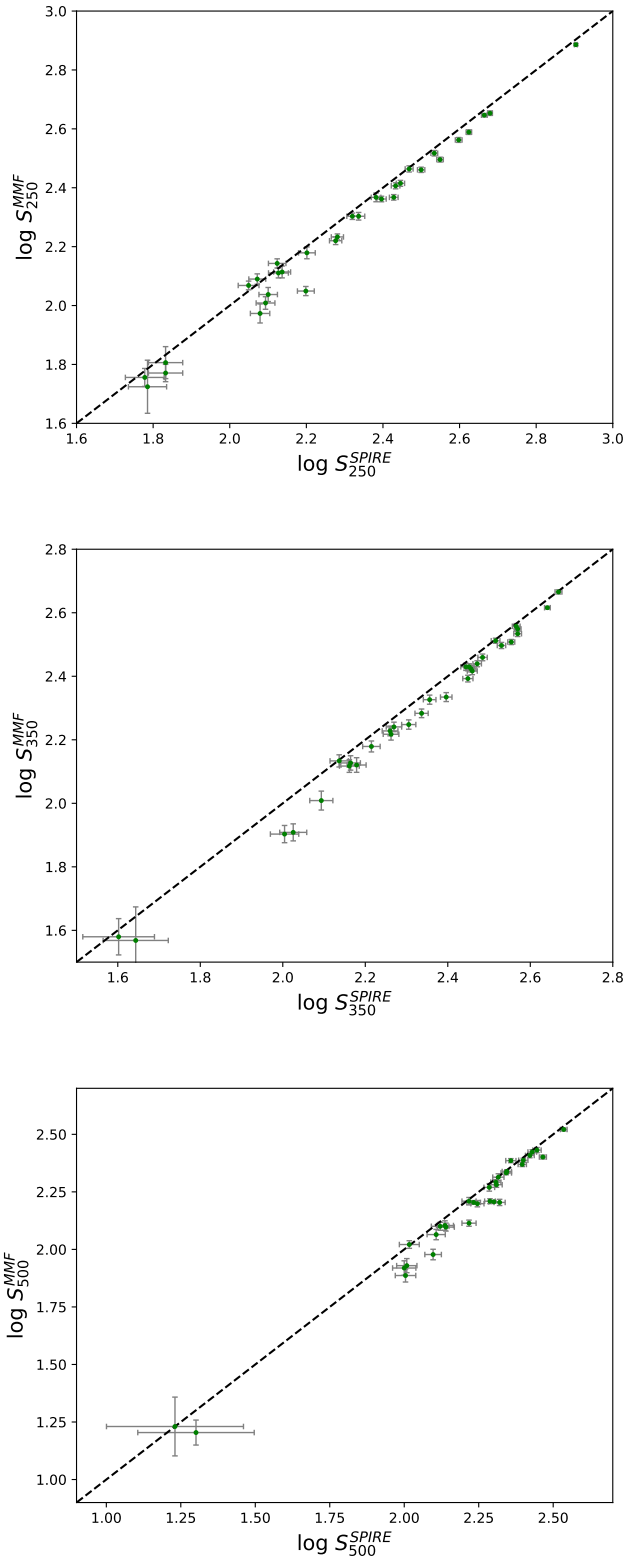


Figure 2.3: Flux measurements in the 250 (top), 350 (middle), and 500 μm (bottom) SPIRE channels derived with the MMF versus the corresponding tabulated H-ATLAS fluxes for the 32 known-spectroscopic sources from Table 2.1. The dashed line marks perfect correlation. All fluxes are in units of mJy/beam, according to the beam profile of the respective channel.

2.6 High-z candidates in H-ATLAS

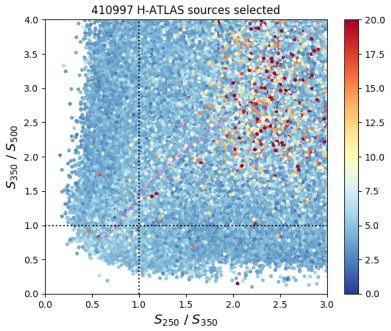
In this section we describe our strategy to find high-z candidates in the H-ATLAS data. In the two subsections below, we explore two different strategies. A first subsample (Sec. 2.6.1) is defined, where the candidates have to be visible in all three SPIRE bands (this will define our bright subsample of high-z candidates) and for which a more reliable estimate of the redshift can be obtained. In the second subsample (Sec. 2.6.2), we focus on the 500-μm risers where the highest flux is found in the band with the longest wavelength. Although this does not guarantee that the source is at high redshift, all of the most distant objects in H-ATLAS will be 500-μm risers, as the peak of the IR emission will be at wavelengths longer than 500-μm.

2.6.1 The bright subsample

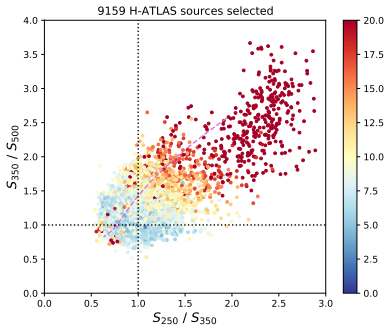
In order to define a sample of reliable high-z candidates in the H-ATLAS, and motivated by the work exposed in Negrello et al. (2017), we follow a strategy based on applying different cuts to the official photometric catalog and the results of running the MMF. This strategy is based on flux cuts, color cuts, and agreement of fit between the photometric measurements and the SED assumed. By applying a series of cuts to the full sample, we reduce the number of candidates until we arrive to a small subsample of objects which meet all our criteria. Since we focus on high-z candidates, some of the cuts are designed to remove low-z sources. We describe these cuts in detail below:

i) Since our method only works for the case of strictly point sources and high-z galaxies will appear as unresolved, first we discard all sources we know for sure are not point-like, that is, sources that have aperture fluxes at the three SPIRE wavelengths different from the point source fluxes. We retain only those sources for which their aperture radius has '-99'-value in the H-ATLAS catalog, which means that the aperture flux and the point source flux are the same. We also remove sources identified as stars and those with null or negative fluxes in any of the channels. This results in a sample of 410,997 objects from H-ATLAS (see Fig. 2.4a), on all of which our MMF method is applied.

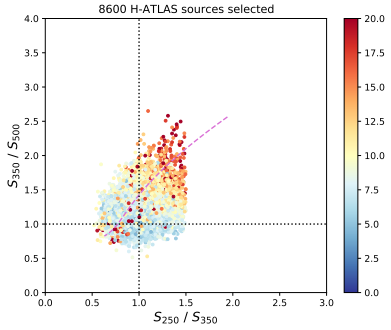
The redshift distribution found for these H-ATLAS objects is shown in Fig. 2.5. An important peak close to zero-redshift can be seen. This figure shows clearly that there are many sources that could be either, a) low redshift sources or, b) sources for which their frequency dependence does not resemble the SED considered. They are not adequate to be used with our method since they would result in erroneous redshifts and fluxes.



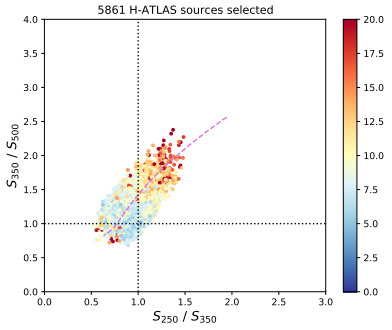
(a) Removal of stars and extended sources.



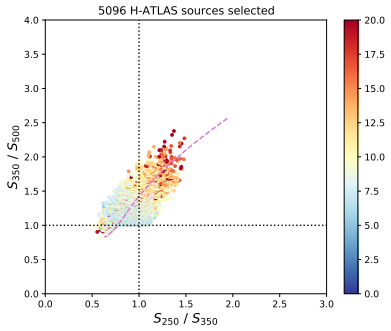
(b) S/N $\geq 5\sigma$ in the three SPIRE channels.



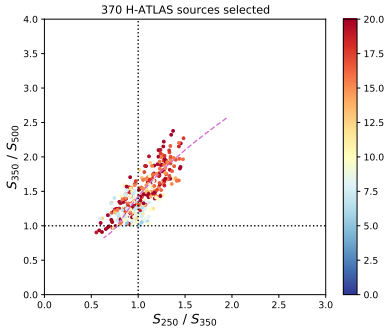
(c) Exclusion of local galaxies at low redshifts.



(d) Closeness to the SED model.



(e) Exclusion of possible blazars.



(f) S/N $\geq 15\sigma$ except if there is an association.

Figure 2.4: Evolution of the color-color diagram of the H-ATLAS sources studied as cuts are applied in order to get a sample of robust high-z candidates. The dashed pink line is the track of the Pearson et al. (2013) SED for redshifts in the range [0.5,4.5] (in increasing order from the top-right to the bottom-left corner). The vertical lateral colored bar present in all plots is a scale of the S/N of the sources exhibited, achieved with the MMF.

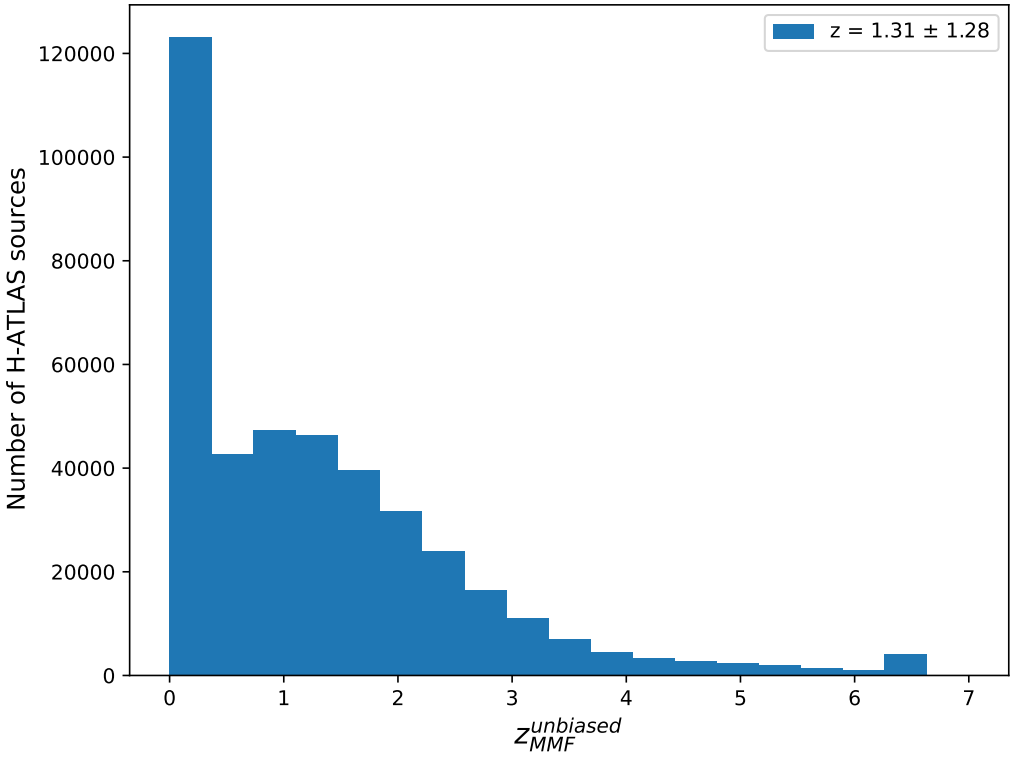


Figure 2.5: Redshift distribution, according to the redshift estimates obtained with the MMF method, of the 410,997 sources from H-ATLAS selected after removing the non point-like objects and those identified as stars.

ii) After removing stars and non-point sources, we proceed to make a preselection using the photometric information of the H-ATLAS catalog. We want our sources to be at high redshift and bright enough so our redshift estimations are robust. Thus we select those which have a S/N greater than 5σ in all three SPIRE channels, which leaves us with just 9,159 sources, 2.2% of the total (see Fig. 2.4b). This is the most stringent cut.

iii) Good candidates are required to have a flux ratio between 250 and 350 μm bands less than or equal to 1.5 (i.e., $S_{250}/S_{350} \leq 1.5$). This cut has the effect of excluding local galaxies at low redshifts. (see Fig. 2.4c).

iv) One important requirement for our preselection is to ensure that the chosen sources have a photometric behavior close to the response offered by the Pearson model (Eq. 2.1) used to estimate their redshifts, since, as we discussed earlier in Sec. 2.2, the method does not work equally well for all H-ATLAS sources (see Fig. 2.6). This can clearly be seen reflected in the large number of sources far away from the Pearson et al. (2013) SED model in Fig. 2.4a. We discard sources that are at a distance larger than 0.3 from the SED model in the color-color diagram (see Fig. 2.4d), according to:

$$\sqrt{\left(\frac{S_{250}^{SPIRE}}{S_{350}^{SPIRE}} - \frac{S_{250}^{MMF}}{S_{350}^{MMF}}\right)^2 + \left(\frac{S_{350}^{SPIRE}}{S_{500}^{SPIRE}} - \frac{S_{350}^{MMF}}{S_{500}^{MMF}}\right)^2} < 0.3 \tag{2.3}$$

The number 0.3 is a compromise between a more stringent requirement that would result in a smaller number of candidates and a more relaxed requirement that would increase the number of candidates but at the expense of increasing the number of sources with unreliable redshift estimations.

v) The last requirement in our preselection is to exclude the presence of possible blazars. As showed in Negrello et al. (2017), the leaking of blazars into a catalog of high-z candidates can be reduced by demanding our sources to have $S_{350}/S_{500} > 1$, unless this ratio is already above the Pearson model in the color-color diagram (see Fig. 2.4e).

At this point, after the cuts i) through v), 5,079 sources remain in the sample. Since this sample will be used later, we denote it the "full high-z" sample. These cuts are not perfect at removing low-z objects but the sample should be dominated by high-z candidates.

vi) Finally, to reduce this level of contamination, and select high-S/N sources for which photometry is expected to be robust (and consequently the photometric redshift as well), we impose the condition that the S/N defined in Eq. (1.8) must be greater than or equal to 15σ in the filtered image after our MMF has been applied, except if the H-ATLAS SPIRE position has an association with a galaxy of known spectroscopic redshift within a separation of 5 arcsec. This cut leaves 370 objects (see Fig. 2.4f). This selection constitutes

our "robust high-z" sample of high-z candidates, with a redshift distribution of $\overline{z} = 2.13$ and $\sigma_z = 0.65$, and will be used later to identify possible lensed systems.

This sample is fully shown in Appendix B, including the estimated redshifts and flux densities. We have performed an additional study for the objects within this sample by comparing the Pearson's χ^2 value obtained considering only the SPIRE flux densities with that obtained taking into account also the PACS flux densities. For this we have used the flux densities from the H-ATLAS catalog as the observed data and the frequency dependence provided by the Pearson SED (Eq. 2.1), at the photometric redshift estimated by our MMF, as the theoretical data. The result of this study is also shown in Appendix B through the flag 'Reliability'. Those sources for which the χ^2 improves or remains the same when using PACS fluxes are flagged with a '0', whether the χ^2 worsens slightly but it is still acceptable they are flagged with a '1', if the χ^2 is much worse, we use '2', and those sources without PACS flux measurements available are flagged with a '-1'.

Within this sample, 201 candidates are in the GAMA fields (60 in the GAMA9, 58 in the GAMA12 and 83 in the GAMA15), 82 in the NGP and 87 in the SGP. The number density of sources in the GAMA fields is higher than in NGP and SGP after this last cut because the number of associations with objects having spectroscopic redshifts is higher in the GAMA fields. 21 out of the 32 spectroscopic redshift sources shown in tables 2.1 to 2.3 (highlighted in bold) are included in this sample. Among all the objects of this sample we find 35 QSOs. Fig. 2.7 shows the redshift distribution of the robust high-z sample in order to compare it with the redshift distribution of the initial sample, shown in Fig. 2.5.

A direct comparison between our estimates of the flux densities in all channels and the tabulated fluxes from the H-ATLAS catalog is shown in Fig. 2.8. A clear linear trend is observed and the agreement is pretty good. As happens with the spectroscopic redshift sources from Fig. 2.3, an overall underestimation of our MMF flux densities, greater for fainter sources, can be observed. As the number of sources is greater, here the effect is most remarkable. The average flux underestimates between the flux densities estimated from the MMF method and the H-ATLAS fluxes are 10 ± 9 mJy at 250 μm, 12 ± 9 mJy at 350 μm and 9 ± 8 mJy at 500 μm. Since the MMF combines information from all three wavelengths, which allows to reduce the background and boost the signal, instrumental, foreground and confusion noises are better removed so flux density estimates are less affected by Eddington bias than H-ATLAS flux densities. This underestimation with respect to H-ATLAS flux densities is stronger toward low flux densities, which supports the Eddington bias hypothesis. However, it is also observed to a lesser extent for high flux densities, suggesting a possible degradation of the MMF photometry that could be related to the way we re-pixelize and combine 350 and 500 μm images in the filtering step (see Sec. 2.3).

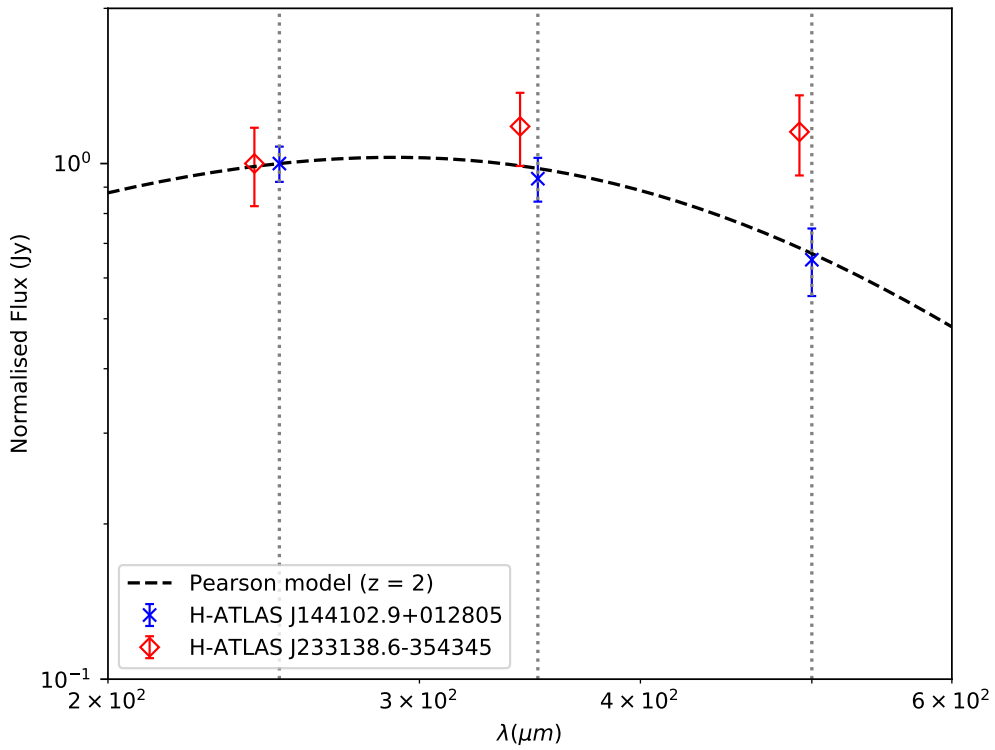


Figure 2.6: Normalized SED from Pearson et al. (2013), as defined in Eq. (2.1), at $z = 2$ in contrast with the normalized tabulated fluxes at 250, 350 and 500 μm (vertical dotted lines) of two H-ATLAS sources at $z_{MMF} = 2$: one (J144102.9+012805) that fits well to the model and other one (J233138.6-354345), whose points have been slightly shifted in the x-axis to get better clarity, that does not fit properly to the model, according to our criteria.

Fig. 2.9 shows the improvement in S/N achieved with our MMF method for the robust high-z sample in contrast with the S/N of the three μm SPIRE bands. An average improvement of 76% in the S/N has been achieved for this sample with our MMF technique compared to the 500 μm band. Besides, an average improvement of 16% and a slight improve of 0.2% have been obtained for the 350 μm and 250 μm, respectively.

In the end, we have a selected sample that includes several hundreds of interesting objects from H-ATLAS, which both agree with the Pearson et al. (2013) SED used to estimate the redshift and have high redshifts and signal-to-noise ratios.

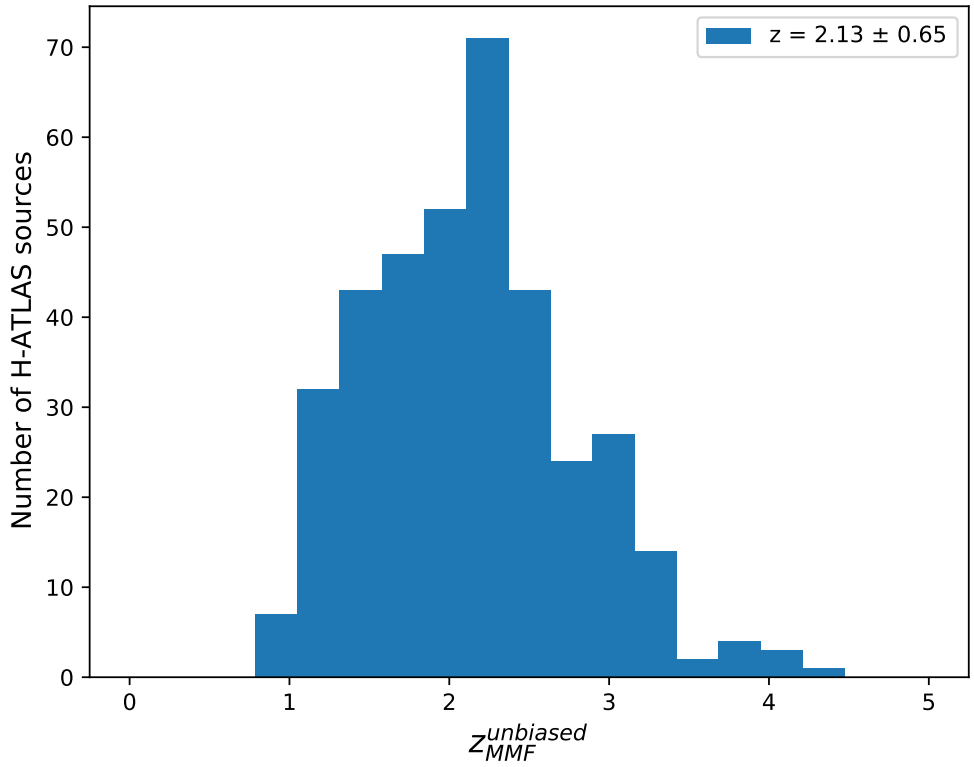


Figure 2.7: Redshift distribution, according to the redshift estimates obtained with the MMF method, of the 370 high-z H-ATLAS sources from the robust high-z sample selected after imposing all cuts enumerated in Sec. 2.6.1.

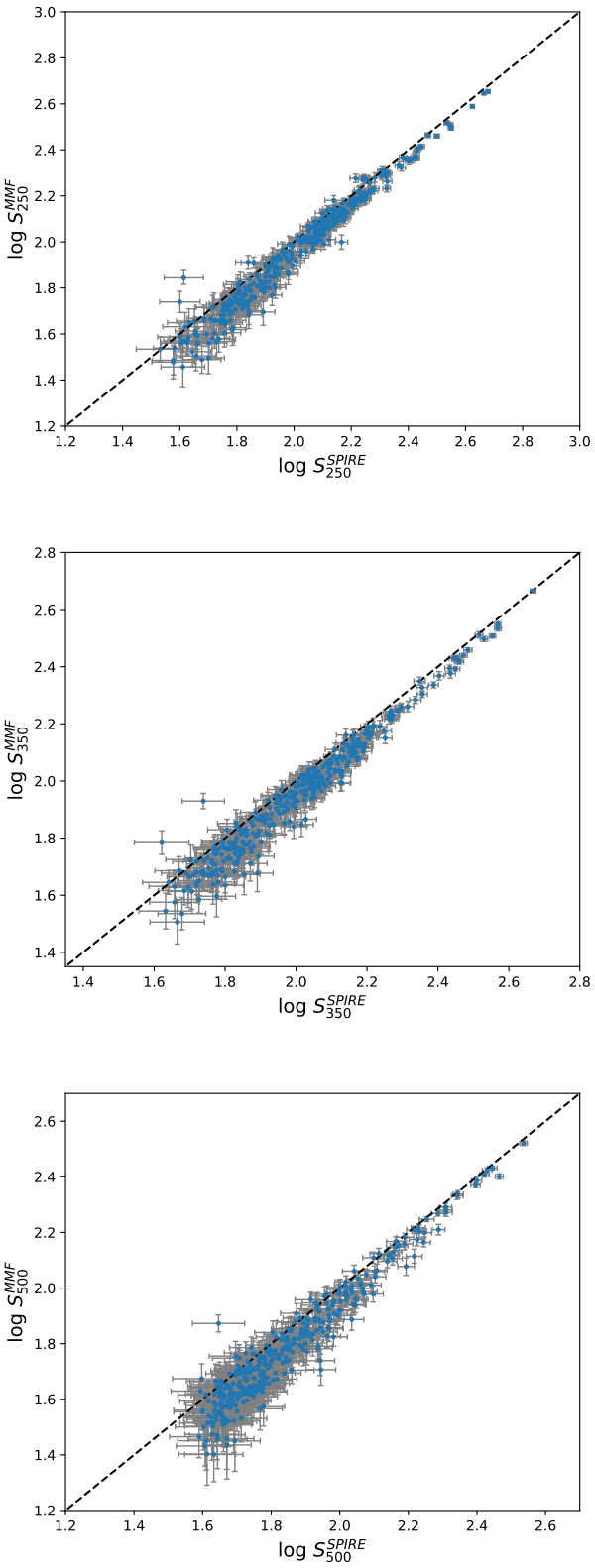


Figure 2.8: Flux measurements in the 250 (top), 350 (middle), and 500 μm (bottom) SPIRE channels derived with the MMF versus the corresponding tabulated H-ATLAS fluxes for the 370 objects in the robust high-z sample. The dashed line marks perfect correlation. All fluxes are in units of mJy/beam, according to the beam profile of the respective channel.

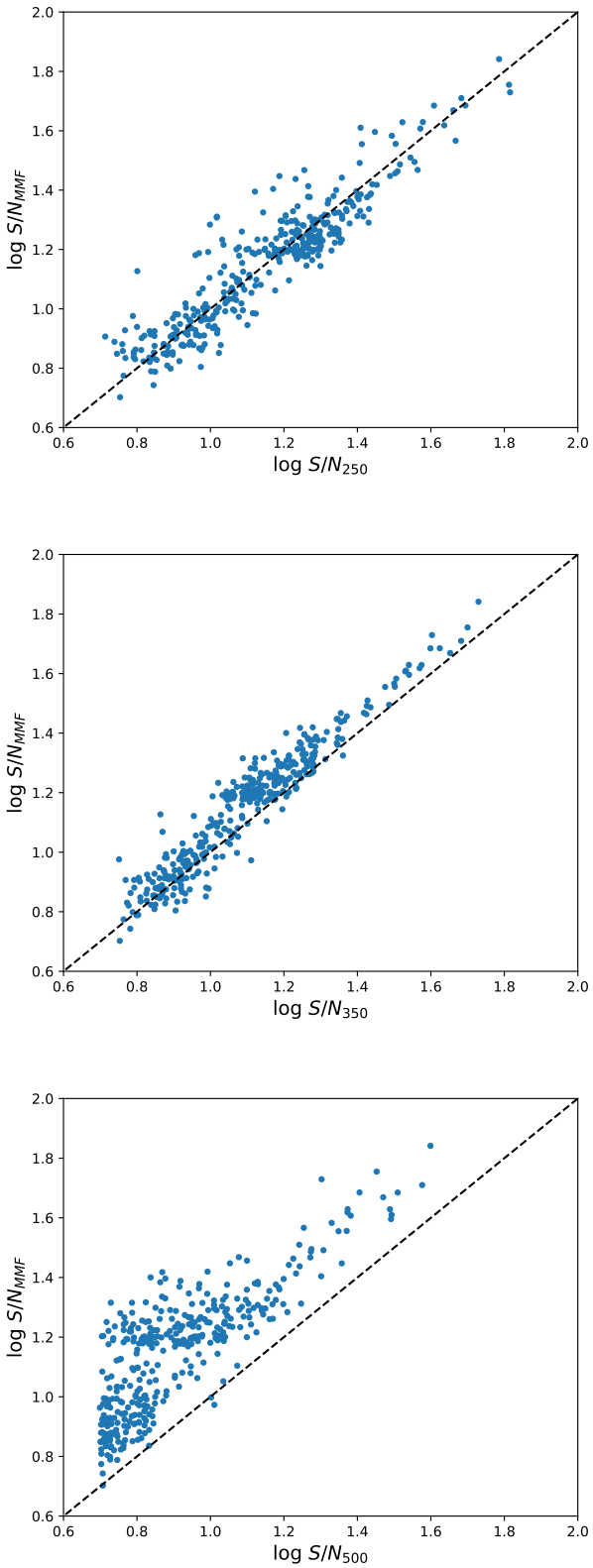


Figure 2.9: S/N in the filtered image derived with the MMF versus the S/Ns in the 250 (top), 350 (middle) and 500 μm (bottom) SPIRE channels for the 370 sources in the robust high-z sample. The linear behavior with zero-intercept is drawn with a black dashed line.

2.6.2 The faint subsample: "500 μm-risers"

Apart from the robust high-z sample explained above in Sec. 2.6.1, we also searched for faint sources at 250 and 350 μm but bright at 500 μm in the H-ATLAS data, the so-called "500 μ-risers". Our selection criteria looks for sources whose detection is at least significant at 500 μm but are not clearly detected at 250 and 350 μm. We select objects with $S/N_{500} \geq 4\sigma$, $S/N_{250} \leq 4\sigma$ and $S/N_{350} \leq 4\sigma$ in the H-ATLAS catalog and apply our multifrequency MMF filter to them, in order to enhance the statistical signification of those detection candidates. Those sources with $S/N \geq 5\sigma$ after the MMF filtering, and that satisfy the condition $S_{250} \leq S_{350} \leq S_{500}$, are considered statistically significant enough to be firm 500 μm-riser candidates. This way, we get a sample of 695 reddened SPIRE objects. However, we must not forget the limitations of the Pearson et al. (2013) SED assumed to estimate the redshifts. Thus, selecting again the sources that best suit the model in the color-color diagram (Eq. 2.3), we are left out with 237 objects. This selection constitutes our 500 μm-riser sample of robust high-z candidates, which have a redshift distribution characterized by $\overline{z} = 4.62$ and $\sigma_z = 0.71$. This entire sample is shown in Appendix B, including the redshift and flux density estimates. The same additional χ^2 study, considering PACS flux densities, performed for the "robust high-z" sample has been applied here, and the result is shown in the appendix through the same flags explained in Sec. 2.6.1. Within this sample, 97 objects are from the GAMA fields (27 in the GAMA9, 37 in the GAMA12 and 33 in the GAMA15), 68 from the NGP and 135 from the SGP.

The comparison between our estimates of the flux densities in all channels and the tabulated fluxes from the H-ATLAS catalog for the 500 μm-riser sample is shown in Fig. 2.10. A much larger scattering than the one seen in Fig. 2.8 for the robust high-z sample can be observed. But this behavior was expected as we are aiming to sources that have a barely significant detection at 500 μm and are not detected at 250 and 350 μm. The average flux underestimates between the flux densities estimated from the MMF method and the H-ATLAS fluxes are 4 ± 4 mJy at 250 μm, 0.4 ± 4 mJy at 350 μm and 3 ± 5 mJy at 500 μm. Fig. 2.11 shows the comparison between the S/N reached with our MMF method and the S/N in all three SPIRE channels. It seems logical that the improvement achieved with our method in S/N for these 500 μm-riser objects (Fig. 2.11) should be better than for the objects from the robust high-z sample (Fig. 2.9), as they are near the H-ATLAS detection limit. This is confirmed since we have achieved average improvements of 25%, 55%, and 76% in the S/N for the 500, 350 and 250 μm, respectively. This clearly reflects that it is in this kind of faint objects where our MMF method accomplishes a bigger impact in terms of signal significance.

Unlike the previous robust high-z sample that sought to select bright objects in all bands, among which the probability of finding lensed systems is relatively high, now we pursue faint high-z objects. Most of them are not expected to be lensed by foreground sources but to be intrinsically luminous. These robust high-z and 500 μm-riser samples are, in fact, built from starting requirements mutually excluding. Sources from the robust high-z sample are initially required to have S/N_{250}, S/N_{350} and S/N_{500} greater than 4σ while sources from the 500 μm-riser sample are demanded to have $S/N_{500} \geq 4\sigma$, $S/N_{250} \leq 4\sigma$, and $S/N_{350} \leq 4\sigma$. But this does not mean, for example, that there are not lensed sources candidates among 500 μm-riser galaxies. It should be pointed out, for instance, the case of J090045.4+004125 ($\alpha = 135.191$, $\delta = 0.6897$), a DSFG at $z = 6$, revealed by strong lensing and detected in the GAMA field as part of a subsample of 500 μm-riser galaxies (Zavala et al., 2018). This object is part of our 500 μm-riser sample under the identifying name J090045.5+004131 with a redshift estimation of $z = 6.35$ via our MMF technique.

An important effort was made recently by Ivison et al. (2016) in order to take advantage of the 250, 350 and 500 μm *Herschel*-ATLAS imaging survey and select extremely red objects. That work focused on studying the space density of luminous DSFGs at $z > 4$ by selecting galaxies from the H-ATLAS survey with extremely red far-infrared colors and faint 350 and 500 μm flux densities, called ultra-red galaxies. It is important to bear in mind that they used a modified version of the MADX algorithm to identify their sources, so some of their sources are not in the official H-ATLAS catalog. This fact explains why we have been able to locate in the H-ATLAS catalog only 78 of the 109 sources shown in their sample. These ultra-red galaxies were required to be above 3.5σ in any of the three SPIRE bands. None of these 78 red sources is in our robust high-z sample. Instead, and as expected, there is an overlap between our 500 μm-riser sample and these red sources. We remind that our 500 μm-riser sample was built demanding $S/N_{500} \geq 4\sigma$, $S/N_{250} \leq 4\sigma$ and $S/N_{350} \leq 4\sigma$. To begin with, 54 out of those 78 objects are strangers to the 500 μm-riser sample because they all have $S/N_{350} > 4$, and thus are excluded by our criterion, which leaves us with 24 possible objects. Only nine of these remaining 24 objects from Ivison et al. (2016) (J090045.5+004131, J090304.5-004616, J114038.8-022804, J114350.3-005210, J114353.5+001250, J114412.1+001812, J115614.0+013900, J142710.6+013806 and J004615.0-321825) are included in our 500 μm-riser sample. If sources were not demanded to behave like the specific SED, their number would increase to 16 objects.

In the next section, we will search for possible lensed systems involving our high-z sources. We will focus on the robust sample because we find no significant correlation for the case of the 500 μm-riser sample. This suggests us that this last sample is mostly not lensed as expected due to the lower flux criterion used to select its sources.

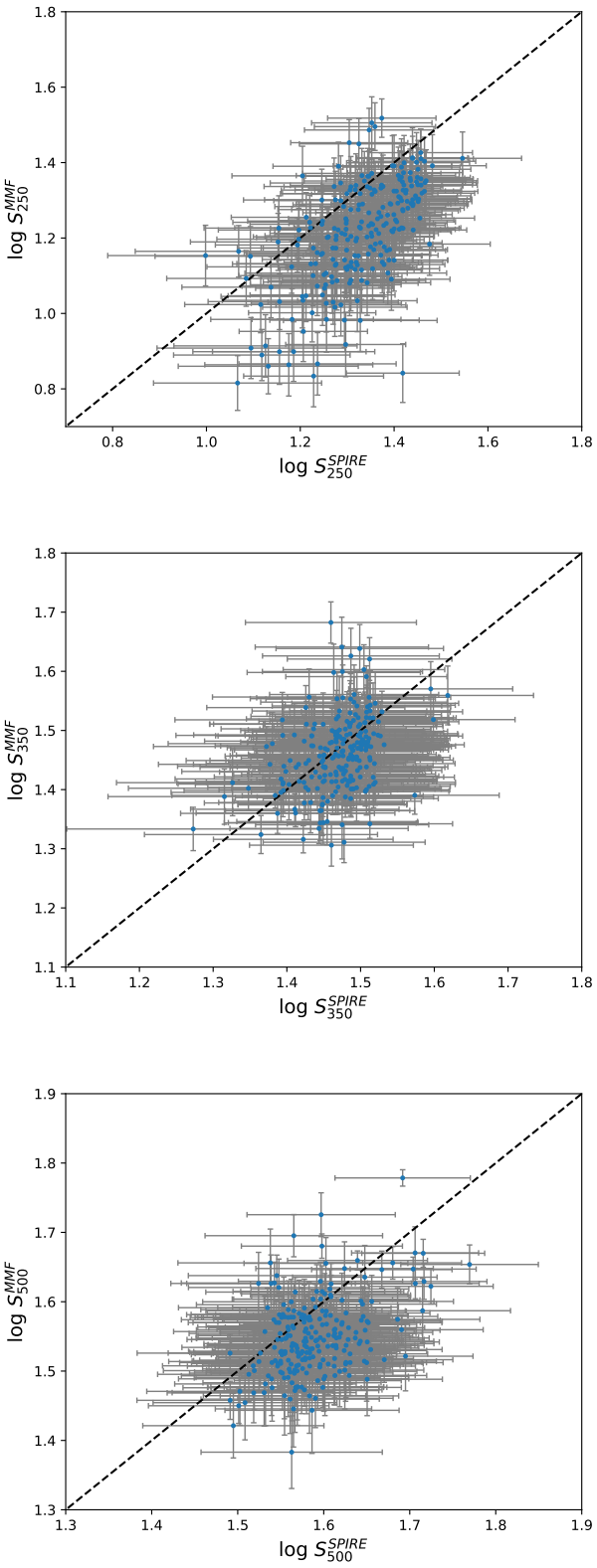


Figure 2.10: Flux measurements in the 250 (top), 350 (middle) and 500 μm (bottom) SPIRE channels derived with the MMF versus the corresponding tabulated H-ATLAS fluxes for the 237 objects in the 500 μm-riser sample. The dashed line marks perfect correlation. All fluxes are in units of mJy/beam, according to the beam profile of the respective channel.

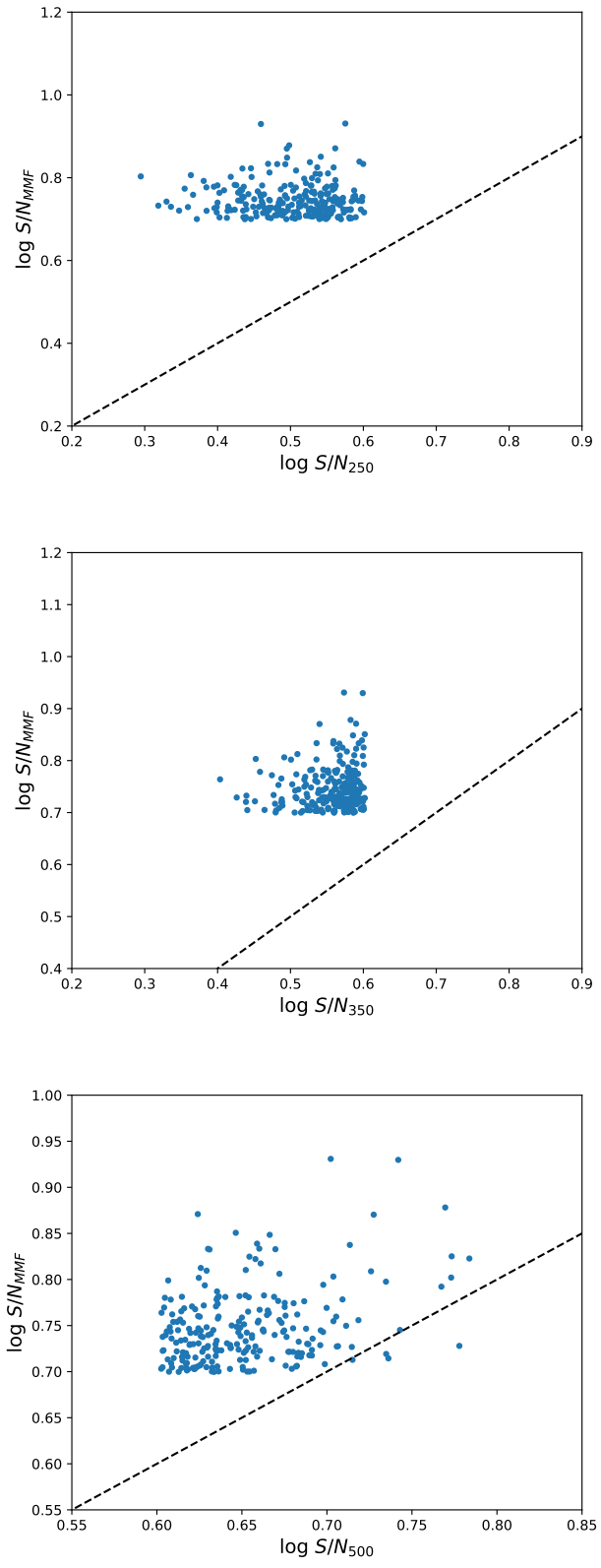


Figure 2.11: S/N in the filtered image derived with the MMF versus the S/Ns in the 250 (top), 350 (middle) and 500 μm (bottom) SPIRE channels for the 237 objects in the 500 μm-riser sample. The dashed line marks perfect correlation.

2.7 Possible lensed galaxies

2.7.1 Preliminary comparison with previous works

In the previous section, we presented our samples of high-redshift candidates (full, robust high-z, and 500 μm-riser samples). Since the optical depth of strong lensing grows with the redshift of the background source, these samples of high-z candidates may contain some lensed galaxies. In fact, this is known to be particularly true for *Herschel* sources, where the brightest high-z sources correspond to SLGs (Negrello et al., 2007, 2010, 2014, 2017; Cox et al., 2011; Conley et al., 2011; Lapi et al., 2011; González-Nuevo et al., 2012; Messias et al., 2014; Dye et al., 2015).

Here we compare the robust high-z sample (370 candidates) with similar catalogs found in the literature. Our robust high-z sample contains 62 of the 80 SLGs candidates with flux densities above 100 mJy at 500 μm presented in Negrello et al. (2017). 17 of the candidates in the robust high-z sample are part of the sample of 20 confirmed SLGs. The only three confirmed strong lens systems that are not included in our sample are: J085358.9+015537 (flagged as a star), J142935.3-002836 (which is a major merger system at $z = 1.027$ (Messias et al., 2014) and is excluded by our cut iii), and J125135.3+261457 (excluded by our cut v). Among the confirmed lensed galaxies, J114637.9-001132 at $z = 3.26$ is interesting since it is associated to a high-z proto-cluster candidate[3](Fu et al., 2012; Herranz et al., 2013; Clements et al., 2016; Greenslade et al., 2018).

In addition, six of the eight objects labeled in Negrello et al. (2017) as likely to be lensed, and 39 of the 51 objects defined as unclear are included in our robust high-z sample of 370 candidates. The two missed objects labeled as likely to be lensed were excluded in our cut iv. The only one object from Negrello et al. (2017) confirmed to not be a strongly lensed galaxy (J084933.4+021442) is nor part of our sample because it is flagged as a star in the H-ATLAS catalog. It is indeed a binary system of Hyper Luminous Infrared Galaxies (HyLIRGs) at $z = 2.410$ (Ivison et al., 2013). Our sample also contains five sources from the SGP field (J004736.0-272951, J011424.0-333614, J235623.1-354119, J001010.5-360237 and J014849.3-331820) which meet the flux criterion demanded by Negrello et al. (2017) but are not in their proposal.

In González-Nuevo et al. (2012), the authors applied to the H-ATLAS Science Demonstration Phase field ($\simeq 14.4$ deg^2), which covers part of the GAMA9 field, a method for efficiently selecting faint SLGs candidates. This method was called HALOS (*Her-*

[3]We have recently submitted an ALMA proposal to confirm the redshifts of the objects associated to this possible proto-cluster.

schel-ATLAS Lensed Objects Selection). They found 31 SLGs candidates, whose respective lens candidates were identified in the VIKING near-infrared catalog, and proposed that the application of HALOS over the full H-ATLAS surveyed area would increase the size of the sample up to $\sim$ 1000 SLGs. Eight of these sources are included in our robust high-z sample: J090302.9-014127, J090311.6+003906, J090740.0-004200, J091043.0-000321, J091304.9-005343 (all of them confirmed as strongly lensed in Negrello et al. 2017), J085855.3+013728, J090957.6-003619 and J091331.4-003644.

The H-ATLAS catalog can be used to find potential lens systems (lens plus lensed galaxy) using the already available optical associations with SDSS (Blanton et al., 2017) for each SPIRE source (Bourne et al., 2016). These associations are sought via a Likelihood-Ratio analysis of optical candidates within 10 arcsec of all SPIRE sources with S/N $\geq$ 4 at 250 μm. Bourne et al. (2014) studied the fact that redder and brighter sub-mm sources have optical associations with greater positional offsets than would be expected if they were due to random positional errors. They concluded that lensing is the most plausible cause for increased offsets of red sub-mm sources and that the problem of misidentifying a galaxy in a lensing structure as the counterpart to a higher redshift submillimeter galaxy may be more common than it was previously thought. Most of these optical associations do not have spectroscopic information (i.e., secure redshift). However, there are 180 objects in our robust high-z sample for which this condition is fulfilled (mostly because of the cut vi). Spectroscopic redshifts are obtained from many different surveys, like SDSS DR7, SDSS DR10, 6dFGS, 2SLAQ or GAMA. 138 sources out of these 180 have a reliable spectroscopic redshift (Z_QUAL $\geq$ 3) in the range $0.1 \leq z \leq 1.1$, which is significantly smaller than the photometric redshift estimated by our MMF method. Hence, these associations may correspond to possible lens systems since the redshifts of the alleged lens and the high-z candidate are so different. In those cases where the optical association is not the same object as the SPIRE source, it will be an object at a smaller redshift and close (in angular separation) to the SPIRE source. The conditions would be given for the lens effect to take place and these cases should be studied in detail to confirm it. However, since these associations are already given in the catalogs themselves and their spectroscopic redshifts come from many different sources, we are going to proceed to look for our own associations.

The above discussion shows how our robust high-z sample has the potential to host many unknown lensed galaxies. Most of the previously known *Herschel* lensed galaxies were unveiled by the 500 μm flux density criterion ($S_{500} > 100$ mJy), which has proven to be a simple, but powerful, method of selecting strongly lensed candidates. Here we rely on a cross-correlation study based on matching distant IR sources with foreground potential lenses located at distances that make them consistent with being a lens system.

2.7.2 Statistical lensing analysis. Correlation analysis with SDSS

Additional evidence for significant lensing in our two samples (full and robust high-z) can be obtained through a simple correlation analysis with a catalog of foreground galaxies. If the *Herschel* sources are tracing the magnification pattern produced by a population of lenses at $z < 1$, one would expect an excess of IR sources detected around regions of magnification larger than one. Alternatively, the alleged high-z source could be instead a lower redshift associated with the lens. In this case, the excess found in the correlation would be produced by contamination of our sample (i.e., low-z sources being misinterpreted as high-z sources).

For the catalog of potential foreground lenses ($z < 1$), we use lenses extracted from the SDSS. By lenses, we mean here either individual galaxies or groups of galaxies (see below). Since SDSS does not cover the SGP field, we consider only the IR sources that come from the GAMA and NGP fields. After removing IR high-z candidates from the SGP field, the full high-z sample is reduced to 2828 candidates while the robust high-z sample is left with 283 candidates. For a simple estimation of the correlation, we compare the number of matches found within an aperture and for different aperture radii, $N_m(R)$, with the expected number from a random distribution ($N_r(R)$, see Eq. 2.4). This random number is obtained by the following equation:

$$N_r\left(R\right) = N_H \frac{A_c\left(R\right)}{A_H} \tag{2.4}$$

where N_H is the number of H-ATLAS high-z candidates, $A_c(R)$ is the total area covered (within the footprint of H-ATLAS) by the disks of radius R around the SDSS sources, and $A_H = 341.65$ square degrees is the total area of H-ATLAS survey excluding the SGP field. By construction, $N_r(R) \leq N_H$.

On the other hand, the number of matches ($N_m(R)$) between the H-ATLAS sources and the SDSS lenses is obtained by computing the number of associations between both catalogs as a function of radius, by centering disks of radius R on the SDSS lenses and counting the number of H-ATLAS sources which fall within the disk. Any significant excess over the expected value in the random case is evidence for either lensing or contamination. The uncertainty, or significance, with respect to the random case is given by the Poissonian error (i.e., the uncertainty is given by $\sigma_r(R) = \sqrt{N_r(R)}$). If the excess is due to contamination, this hypothesis can be tested, since one would expect the separation between the positions of the *Herschel* sources and the SDSS lenses to be comparable to the positional error in *Herschel* (which is significantly larger than the corresponding error in SDSS), that is

2–3 arcsec. In these cases, the *Herschel* source may actually be the SDSS lens. If, on the contrary, a high-z *Herschel* candidate is found at more than 3 arcsec from the SDSS source, lensing is possibly responsible for that association. Some of the associations should be due to pure random alignments, but this number can be estimated by the Poissonian expectation number discussed above.

For the SDSS lenses, we use two catalogs of potential lenses derived from SDSS. The first catalog focuses on rare but massive potential lenses at $z \leq 0.6$, while the second catalog focuses on less massive, but more abundant, potential lenses with $z \leq 1.1$. We set a lower limit to the redshift of the potential lenses since below certain redshift, strong lensing becomes inefficient due to the increase in the critical surface mass density ($z_{min} \sim 0.1$).

For the association with massive lenses, we use the SDSS DR8 redMaPPer cluster catalog with 26111 objects (Rykoff et al., 2014). This catalog is the result of applying the Red-sequence Matched-filter Probabilistic Percolation (redMaPPer) cluster finding algorithm to the 10400 deg^2 of photometric data from the Eighth Data Release (DR8, SDSS-III collaboration, 2011) of the SDSS. The redMaPPer algorithm has been designed to handle an arbitrary photometric galaxy catalog, with an arbitrary number of photometric bands (≥ 3), and performs well provided the photometric bands span the 4000 Å break over the redshift range of interest. It adapts therefore well to a survey such as the Sloan Digital Sky Survey. Because the number of objects in this catalog is not very large, we use all of them in the cross-correlation which cover a range of $0.08 \leq z \leq 0.6$ in redshift and $19.85 \leq \lambda \leq 299.46$ in cluster richness. $N_H = 881$ of the 26111 halos fall in the footprint of H-ATLAS. We find no significant excess when cross-correlating redMaPPer with our catalog of high-z H-ATLAS sources. Given the fact that 17 of the 20 strongly lensed candidates from Negrello et al. (2017) (confirmed as such) are in our selected sample, and none of them has a match with redMaPPer, this confirms that the lenses in Negrello et al. (2017) are not massive halos, but rather relatively small halos (like elliptical galaxies for instance).

Our second search for potential lensed galaxies focuses on the low-mass regime of the lenses. From SDSS DR14 (Abolfathi et al., 2018) we select a larger catalog of galaxies with known spectroscopic redshifts. We focus on galaxies with known redshift in order to minimize possible contamination from galaxies that are misinterpreted as having $z > 0.15$ and also to reduce the computation time. The sample contains 1,776,242 galaxies from the Sloan Digital Sky Survey Data Release 14 with $0.15 \leq z \leq 1.1$. As mentioned above, we limit the minimum redshift to 0.15 since below this redshift most galaxies are expected to be subcritical (and not produce strong lensing effects). Among all of them, $N_H = 50175$ are the galaxies that fall in the footprint of H-ATLAS. We cross-correlate our full high-z subsample of 2828 H-ATLAS sources with the SDSS catalog of galaxies and compare it

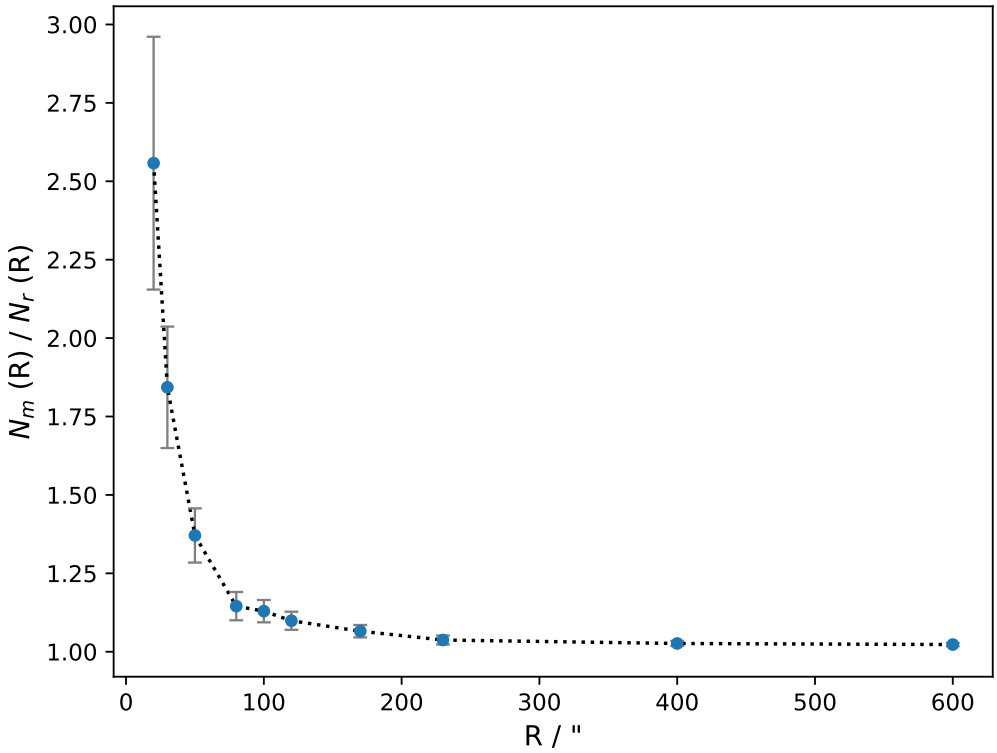


Figure 2.12: Ratio between the number of matches found ($N_m(R)$), after cross-matching the full high-z subsample of 2828 H-ATLAS sources with a sample of 1,776,242 known-redshift galaxies from SDSS DR14, and the number of matches expected ($N_r(R)$) from a random distribution (Eq. 2.4) for different aperture radii R around the SDSS sources.

with the expected number in the case of no correlation (i.e., the random case described above). The ratio of the observed ($N_m(R)$) and random matches ($N_r(R)$) between this catalog and our full high-z selection sample is shown in Fig. 2.12 for different radii. In Fig. 2.13 we exhibit the same but for our robust high-z subsample of 283 candidates. Both Fig. 2.12 and Fig. 2.13 show a non-one signal for aperture radii of several arcminutes, which is unexpected and an example of the lensing-induced cross-correlations between high-z submillimeter galaxies and low-z galaxy population (Wang et al., 2011; González-Nuevo et al., 2014, 2017; Bourne et al., 2014).

There is a clear increase in significance when considering the robust high-z subsample of 283 candidates. A sharp increase in the excess of matches is found at distances below 60 arcsec. The smaller amplitude of the excess in the full high-z sample with 2828 sources suggests that this sample may be more contaminated by low-z candidates.

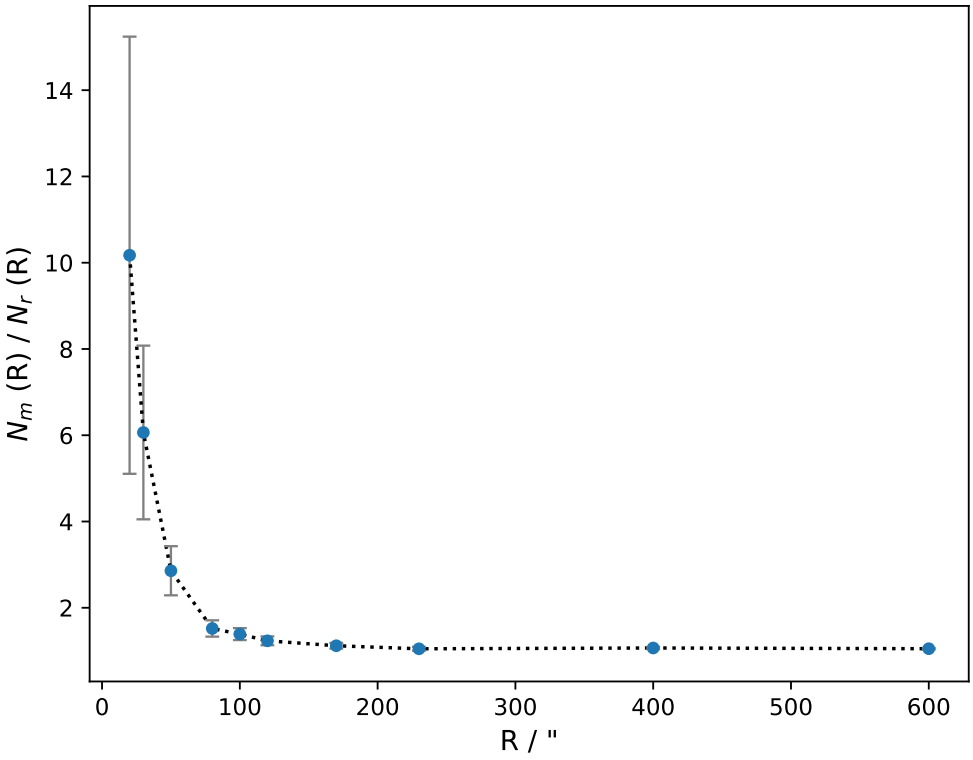


Figure 2.13: Ratio between the number of matches found ($N_m(R)$), after cross-matching the robust high-z subsample of 283 candidates with a sample of 1,776,242 known-redshift galaxies from SDSS DR14, and the number of matches expected ($N_r(R)$) from a random distribution (Eq. 2.4) for different aperture radii R around the SDSS sources.

No.	H-ATLAS name	α (deg)	δ (deg)	$z_{MMF}^{unbiased}$	$z_{spec}^{H-ATLAS}$	SDSS ID	α (deg)	δ (deg)	z_{spec}^{SDSS}	R (arcsec)	θ_E (arcsec)
1	J083726.1+015641	129.358797	1.944871	1.86 ± 0.28	0.396	1237653622837216254	129.359078	1.944271	0.396	2.4	3.9
2	J085229.0+010217	133.121072	1.038117	2.46 ± 0.30	0.492	1237650797288031084	133.120142	1.038521	0.492	3.7	6.9
3	J085250.9-010000	133.21219	-1.000204	1.13 ± 0.26	0.581	1237673709866321335	133.213142	-1.000520	0.583	3.6	2.3
4	J085855.4+013729	134.731174	1.624899	2.36 ± 0.30	0.665	1237651753990489111	134.729564	1.626498	0.658	8.1	26.6
5	J090038.0-003522	135.158643	-0.589715	1.54 ± 0.27	0.407	1237674460412707797	135.154630	-0.586598	0.570	18.3	4.5
6	J090311.6+003907	135.798682	0.652064	3.17 ± 0.34	0.300	1237648722285101574	135.798225	0.651815	0.300	1.9	1.4
7	J091331.4-003644	138.380891	-0.612341	1.49 ± 0.27	0.296	1237648720675668446	138.383247	-0.610143	0.499	11.6	3.4
8	J092348.4+011907	140.952056	1.318877	1.44 ± 0.27	0.590	1237653664181126111	140.952085	1.319323	0.590	1.6	5.4
9*	J114512.9-002446	176.303787	-0.412797	3.13 ± 0.34	0.444	1237671140943987623	176.308407	-0.412486	0.629	16.7	5.6
10	J115101.7-020024	177.757213	-2.006799	1.89 ± 0.28	0.264	1237650762389586324	177.760981	-2.006413	0.454	13.6	6.3
11	J115819.5-000127	179.581642	-0.024356	2.32 ± 0.29	0.597	1237648721230561969	179.579745	-0.026204	0.597	9.5	6.8
12	J120656.4+012641	181.735018	1.444872	2.11 ± 0.28	0.637	1237651752400454138	181.733129	1.443833	0.637	7.8	5.6
13	J120932.7+002517	182.386595	0.421524	1.51 ± 0.27	0.470	1237674651003454270	182.385894	0.420998	0.471	3.2	5.4
14	J125233.5+331031	193.139528	33.175366	2.05 ± 0.28	0.490	1237665331465945161	193.139024	33.175429	0.490	1.5	4.8
15	J130138.7+302808	195.411281	30.468764	2.03 ± 0.28	0.166	1237665443126575176	195.410385	30.468297	0.166	3.3	2.3
16	J130152.8+245012	195.469976	24.836605	2.28 ± 0.29	0.519	1237667911671153185	195.469452	24.836544	0.519	1.7	3.7
17*	J131407.0+271208	198.529016	27.202121	0.96 ± 0.27	0.189	1237667443513229416	198.529224	27.203110	0.189	3.6	1.3
18*	J131533.9+233254	198.891415	23.54824	1.35 ± 0.26	-	1237667910598787245	198.892569	23.550384	0.551	8.6	4.2
19	J131635.1+332406	199.14629	33.401613	3.33 ± 0.35	0.463	1237665127475380722	199.147629	33.401386	0.463	4.1	5.6
20*	J131715.6+322643	199.314904	32.44538	1.25 ± 0.26	0.185	1237665227847565637	199.313765	32.446109	0.185	4.3	11.7
21	J132111.2+265009	200.296539	26.835849	2.08 ± 0.28	-	1237667323262992805	200.293255	26.836073	0.542	10.6	0.6
22	J132355.2+282319	200.979945	28.388498	2.33 ± 0.29	0.406	1237665440981254703	200.978406	28.389180	0.406	5.5	3.1
23	J132453.6+244909	201.223168	24.819245	1.82 ± 0.27	0.569	1237667448346903078	201.221548	24.818992	0.569	5.4	3.2
24	J133020.9+240249	202.587121	24.046861	3.00 ± 0.33	0.610	1237667911673971962	202.586077	24.046954	0.610	3.5	4.2
25*	J133231.6+350843	203.131806	35.145221	1.82 ± 0.27	0.189	1237664852029079667	203.131680	35.145198	0.189	0.4	2.0
26	J133806.8+351530	204.528535	35.258382	1.90 ± 0.28	0.339	1237664852566409566	204.528530	35.257533	0.339	3.1	0.6
27	J134324.5+240202	205.852117	24.034017	2.15 ± 0.29	0.500	1237667912212153027	205.854422	24.035555	0.500	9.4	2.6
28	J134429.5+303034	206.12285	30.509484	2.29 ± 0.29	-	1237665329860248195	206.122041	30.504826	0.672	17.0	0.9
29*	J134853.0+270011	207.220753	27.002946	2.49 ± 0.30	0.175	1237665533319643218	207.223273	27.003177	0.175	8.1	3.0
30	J141351.9-000026	213.466627	-0.007454	2.72 ± 0.31	-	1237674603753243273	213.466488	-0.009563	0.548	7.6	1.2
31	J141550.4+012750	213.960027	1.464006	2.47 ± 0.30	0.296	1237651752951415050	213.959966	1.465131	0.296	4.1	1.5
32*	J141605.5+011828	214.023078	1.308023	3.15 ± 0.34	0.588	1237651735235003221	214.024281	1.307622	0.588	4.6	3.3
33	J141827.4-001703	214.614453	-0.284336	1.85 ± 0.28	0.250	1237648704053903588	214.614920	-0.284529	0.250	1.8	3.6
34	J141832.9+010212	214.637488	1.036847	4.14 ± 0.39	0.668	1237651752414871842	214.636719	1.036710	0.668	2.8	3.2
35	J142008.9-001434	215.03741	-0.242847	2.27 ± 0.29	0.615	1237648704054035361	215.037730	-0.242879	0.615	1.2	5.9
36	J142233.9+023413	215.641415	2.570521	2.20 ± 0.29	0.747	1237651736846270820	215.641393	2.568986	0.747	5.5	0.2
37	J143358.4-012718	218.493585	-1.455029	1.89 ± 0.28	0.239	1237655693012369781	218.49700	-1.456681	0.238	13.7	2.1
38	J143845.8+013503	219.691249	1.584351	1.62 ± 0.27	0.498	1237651735774364011	219.691389	1.583848	0.498	1.9	5.3
39	J145420.6-005203	223.586163	-0.867644	2.58 ± 0.31	0.765	1237648720176087748	223.586516	-0.872245	0.551	16.6	8.4
40	J145653.4-000720	224.222595	-0.122431	1.58 ± 0.27	0.647	1237648721250092410	224.222216	-0.122965	0.647	2.4	1.5

Table 2.4: 40 matches found at a separation radius less than or equal to 20 arcsec after cross-matching the robust high-z subsample of 283 high-z candidates with a sample of 1,776,242 known-low-redshift galaxies from SDSS DR14. The object J145420.6-005203 is identified as a QSO in the H-ATLAS catalog. For each H-ATLAS source the following information is provided: our redshift estimate after revising the bias ($z_{MMF}^{unbiased}$), spectroscopic redshift of the optical association assigned in the H-ATLAS catalog ($z_{spec}^{H-ATLAS}$), identifying name (SDSS ID) of the association found by us, its spectroscopic redshift (z_{spec}^{SDDS}), its separation radius in arcsec (R) and a rough estimation of Einstein radius that the lensed system would have according to the calculations and assumptions described in Sec. 2.7.2. Those sources marked with an asterisk (*) are also forecasted to be lens systems according to the best deep learning model derived from KiDS-like simulations in Sec. 6.1.

Focusing on the smaller radii, we find 40 associations at a separation lower than 20 arcsec, between the 50175 known-spectroscopic-redshift galaxies from SDSS DR14 that fall in H-ATLAS footprint and our robust high-z subsample of 283 H-ATLAS sources. One of them (J145420.6-005203) is identified as a QSO in the H-ATLAS catalog. We chose to consider this separation since most of the lenses would have an Einstein radius less than 20 arcsec, which is the radius around which the strongest magnifications are expected. From among these matches, 28 have a separation greater than the positional error in *Herschel* (>3 arcsec) so lensing is possibly responsible for that association. ~ 4 associations should be caused due to pure random alignments, so it is expected that a considerable number of these associations are lensed. These 40 matches are shown in Table 2.4. And snapshots of them, centered on the SDSS DR14 sources, are shown in Fig. C from Appendix C.

We have used the SDSS DR14 asinh magnitudes in the r-band of these 40 low redshift optical sources shown in Table 2.4 to get a rough estimation of the Einstein radius of each possible lensed system. Firstly, we have estimated from van Uitert et al. (2015) the corresponding corrections for the redshift of their spectra (i.e., the k-correction) and for the intrinsic evolution of their luminosity (i.e., the e-correction) in order to correct the r-band magnitudes. Once the magnitudes of the optical sources are corrected, we have calculated their fluxes and then their luminosities (through their luminosity distances D_L). At this point, we used the luminosity-to-halo mass relation $M_{eff} = M_{0,L}(L/L_0)^{\beta_L}$ parametrized in van Uitert et al. (2015) to estimate the mass of each SDSS galaxy for the corresponding luminosity previously obtained. The pivot luminosity L_0 is the same for every object while the $M_{0,L}$ and β_L parameters depend on the spectroscopic redshift of the galaxy. Finally, we supposed the galaxy behaves as a singular isothermal sphere to estimate the Einstein radius (see Narayan and Bartelmann (1996)). We assumed the virial radius of the galaxy to be $r = 1.3(M/10^{15}M_{\odot})^{1/3}$ Mpc in order to estimate its velocity dispersion $\sigma_v = \sqrt{GM/2r}$. The Einstein radius can be then estimated through $\theta_E = (4\pi\sigma_v^2 D_{ds})/(c^2 D_s)$, where D_{ds} and D_s are the angular diameter distances between the lens and the source, and observer and source, respectively. These distances are calculated with the spectroscopic redshift of the SDSS galaxy acting as lens (z_{spec}^{SDSS}) and the photometric redshift of the source estimated with our MMF ($z_{MMF}^{unbiased}$). These Einstein radius estimates are included in Table 2.4 and endorse the choice of 20 arcsec as upper limit for the separation between H-ATLAS sources and their partners from SDSS DR14. 15 of the 40 matches have a separation radius smaller than the corresponding estimation of the Einstein radius. For these 15 we will expect to find counter-images outside the Einstein radius. Possible reasons why we do not find those could be: i) the foreground galaxy is not a lens, ii) there is substructure around the lens that can create fluctuations in the flux or iii) the Einstein radius is overestimated. Einstein rings corresponding to these radii have been drawn in the snapshots from Appendix C.

All H-ATLAS objects from Table 2.4, except four of them, had a previous association with a galaxy with spectroscopic redshift at less than 10 arcsec. In our search for possible lens systems, we extend our association radius to 20 arcsec. We find that 2 of the previously unmatched H-ATLAS sources (J132111.2+265009 and J134429.5+303034) can now be matched with a galaxy having spectroscopic redshift within 20 arcsec. For the other two remaining unmatched sources (J131533.9+233254 and J141351.9-000026), we find associations within 10 arcsec with SDSS galaxies having spectroscopic redshift. These last two associations may not have been included in the H-ATLAS catalog because SDSS DR7 and SDSS DR10 were used and we cross-match our robust high-z sample with known-low-redshift galaxies from SDSS DR14.

There are five sources for which there is a mismatch between the redshift published in the H-ATLAS catalog and the redshift of the match found in this work. These sources are J090038.0-003522, J091331.4-003644 (SLG candidate from González-Nuevo et al. (2012)), J114512.9-002446, J115101.7-020024 (labeled as unclear in Negrello et al. (2017)) and J145420.6-005203. In all cases, our matches are beyond the 10 arcsec radius used in the H-ATLAS association. All these five sources lie in the GAMA fields for which more redshift information is available other than the SDSS redshifts (and that was not used in our association). In fact, we can confirm that spectroscopic redshifts for the first four sources come from the GAMA project (Driver et al., 2009, 2016) but the last one comes from the WiggleZ Dark Energy Survey.

After visual inspection, among the 40 matches we identify interesting cases that are likely to be lensed. Three of them (J090311.6+003907, J134429.5+303034 and J141351.9-000026) are already confirmed as strongly lensed systems in Negrello et al. (2017). Some other SDSS stamps reveal overdensities of galaxies at $z < 1$ that could indicate the presence of a group of galaxies acting as a lens for the corresponding candidate. These are mainly J083726.1+015641, J085229.0+010217, J085250.9-010000, J115819.5-000127, J130138.7+302808, J131407.0+271208, J131533.9+233254, J133231.6+350843, J133806.8+351530, J141832.9+010212 and J143845.8+013503.

The properties and complex nature of the strong gravitationally lensed system J090311.6+003907 (proposed as SLG candidates in González-Nuevo et al. (2012)) have been studied in detail in Dye et al. (2015) by modeling ALMA long baseline imaging. This source has an optical association in the H-ATLAS catalog at 1.87 arcsec with z=0.3 which is compatible with both the assocciation we found after the cross-matching and the association identified in Negrello et al. (2017).

The source J134429.5+303034 has no association in the H-ATLAS catalog (within 10 arcsec) with an optical counterpart with known-redshift. Although we find an SDSS galaxy

(with unknown redshift) at 0.43 arcsec distance. At larger radii, we find an association with a SDSS galaxy having $z_{spec} = 0.67$ at 16.96 arcsec. In Negrello et al. (2017) they find a potential lens with $z_{spec} = 0.6721$ but much closer to the H-ATLAS source (0.43 arcsec). A closer look at the SDSS images reveals how the potential lens can be barely resolved in SDSS images. We propose that the small lens identified in Negrello et al. (2017) is part of a larger halo (or lens) at $z = 0.67$.

The source J141351.9-000026 resembles the situation of the previous source. It has no optical association with any galaxy having known spectroscopic redshift in the H-ATLAS catalog, but we find a galaxy (with unknown redshift) in the SDSS images at 1.12 arcsec distance. For this source we find an association with a SDSS galaxy at 7.61 arcsec at $z = 0.547$. This association differs from the one in Negrello et al. (2017) which finds a much fainter galaxy at 1.12 arcsec (also at $z = 0.547$). A closer inspection of the optical high resolution image in Negrello et al. (2017) reveals a giant arc at the position of the IR source. The morphology of the arc (with a radius of order 10" and aligned in a tangential direction with respect to our lens candidate) suggests that our association is the correct one. However, the galaxy found in Negrello et al. (2017) could still be contributing to the magnification, since it seems to lie near the critical curve (for the redshifts of the system).

The object J115101.7-020024 is also present in Negrello et al. (2017) proposal. It is labeled as an unclear case because no optical association was found. But we have found an association for this source with $z_{spec} = 0.4543$ at 13.63 arcsec, which differs from the optical association assigned in H-ATLAS catalog.

J131533.9+233254 is an interesting example. The H-ATLAS catalog contains no optical association. We find an association in SDSS at 8.61 arcsec and with $z = 0.551$. The snapshot of this association in Fig. C reveals a number of galaxies in the vicinity of the SDSS galaxy, suggesting that the lens may be a group or a cluster at z=0.55. The stamp for J141351.9-000026 (discussed above) resembles this situation, although with fewer SDSS galaxies. Other interesting candidates are J083726.1+015641, J085229.0+010227 and J115819.5-000127, which show a concentration of SDSS galaxies near the IR source. The already mentioned object J091331.4-003644 and J085855.4+013729 are the only ones candidates from González-Nuevo et al. (2012) presented in the robust high-z sample for which we found associations with spectroscopic redshifts.

The selection shown in Appendix C contains only a small sample of potential candidates to be lensing systems. Our search for this type of systems is limited by the fact that we restrict ourselves to SDSS galaxies with spectroscopic redshifts. Future surveys with many more galaxies having spectroscopic redshifts (like Euclid), or accurate photometric redshifts (like J-PAS) can significantly increase the number of lensed systems.

2.8 Conclusions

In this chapter we have extended previous work on the multifrequency matched multifilter (MMF) technique to sub-mm astronomy, in particular the *Herschel*-ATLAS fields. Our multifrequency technique allows us to, on the one hand, boost the statistical significance and S/N of the detections and, on the other hand, provide an estimation of the photometric redshift. This approach relies on semi-analytic modeling of the thermal peak of the SED of dusty galaxies (see e.g., Pearson et al., 2013). We remark that the use of SED templates is not strictly necessary for the MMF technique[4], but it becomes very useful for the purpose of redshift estimation. We have tested our technique with both realistic simulations and also with H-ATLAS sources for which their redshifts are known, concluding that the MMF indeed leads to an improved S/N with respect to single-frequency detection and that the photometric redshifts are relatively accurate, especially for those H-ATLAS sources with a photometric behavior close the SED used in the range $0.8 < z < 4.3$. Outside of this range, we have studied the bias in the MMF photo-z estimation and provided correction factors up to $z \sim 7.0$.

410,997 point-like sources from the DR1 and DR2 H-ATLAS fields (Valiante et al., 2016; Maddox et al., 2018) have been re-analyzed with this multifrequency filter. Two different sets of selection criteria based on flux, color, S/N, and agreement between observed photometry and the SED model have been carried out in order to select the H-ATLAS sources for which the MMF method performs the best and illustrate its potential usefulness. A total of 607 H-ATLAS sources are found, for which new photometry and photometric redshifts are provided. They are split into two samples of interest:

- A sample of 370 bright high-z objects with $\bar{z} = 2.13$ and $\sigma_z = 0.65$. This sample contains 62 of the 80 strongly lensed candidates identified in Negrello et al. (2017), and including 17 out of the 20 confirmed strongly lensed galaxies. An average improvement of 76% in the S/N has been achieved for this sample with our MMF method compared to the 500-μm band. Besides, an average improvement of 16% and a slight improve of 0.2% have been obtained for the 350-μm and 250-μm, respectively. By cross-correlating this high-z sample with a catalog of low-known-redshift galaxies from SDSS DR14, we find 40 potential candidates to being gravitational lens systems which have been shown and discussed. Some SDSS stamps around these candidates reveal overdensities of galaxies at $z < 1$ that could indicate the presence of a group of galaxies acting as a lens.

[4] As discussed in Lanz et al. (2010, 2013), the MMF can work with arbitrary SED points taken as free random variables, but some physical modeling is often both justified and useful.

- A second sample of 237 faint high significance candidates to 500-μm risers (with $\bar{z} = 4.62$ and $\sigma_z = 0.71$) which were previously near the H-ATLAS detection limit but are now confirmed with the MMF as high significance detections. We have achieved average improvements of 25%, 55% and 76% in the S/N for the 500-μm, 350-μm and 250-μm, respectively. This clearly reflects that it is in this kind of faint objects where our MMF method reaches its full potential in terms of S/N improvement.

The MMF has proven to be able to enhance the sensitivity of fainter objects and improve its photometry. According to the comparative analysis of the photometric redshift estimates with known-redshift H-ATLAS sources, it may also return more robust photometric redshifts than the ones that would be obtained using the fluxes of the H-ATLAS catalog, as proven by the mean μ and median $\mu_{1/2}$ values shown in tables 2.1 and 2.3.

Running a non-blind search, using the already-known source positions in the public H-ATLAS catalog, allows us to directly compare our method with the detection pipeline of the public H-ATLAS catalog. However, given that, in the end, H-ATLAS detection image simply reduces to the 250 μm channel (Valiante et al., 2016), in the future we plan to apply this method in a blind search mode over the H-ATLAS maps. Thereby we will be able to exploit the information from the other wavelengths and be able to detect faint sources that we could be missing now.

The work done here could also be extended in the future focusing only in the sources already confirmed. One of the possible deeper analysis or improvements would be to repeat the MMF execution using other SEDs different from the Pearson et al. (2013) model. For instance, SEDs derived from ultraluminous dusty galaxies Arp200; SMM J2135-0102, "The Cosmic Eyelash" at $z = 2.3$ (Ivison et al., 2010) or H-ATLAS J142413.9+022304 alias G15.141 at $z = 4.23$ (Cox et al., 2011), which are characteristic of the local submillimeter galaxies. This procedure would show us how much the redshifts and flux densities change depending on the SED used and would help us to select sources in a more robust way. This further analysis would also allow us to check whether the methodology bias observed in simulations is because of the particular shape of the function employed. Nevertheless, we think that the Pearson et al. (2013) model is the best model choice since it is not a certain SED from a single source but an empirical model based on 40 H-ATLAS sources and it has already been used in several previous works (Eales, 2015; Ivison et al., 2016; Bianchini et al., 2016, 2018; Negrello et al., 2017; Fudamoto et al., 2017; Bakx et al., 2018; Donevski et al., 2018).

Chapter 3

Gravitational Lensing

3.1 Introduction

The first written reference to the concept of gravitational lensing dates back to 1783. That year, John Mitchell, an unsung English natural phylosopher, sent to Henry Cavendish a letter proposing the idea of light deflection by mass, and describing a method to measure the mass of stars by detecting the reduction in the light speed that this phenomenon causes. Around that time, this suspicion was shared by many others such as Newton, Laplace, or Soldner. It took until the XXth century for more light to be shed on this matter. One of the predictions of Einstein's Theory of General Relativity is that light follows the null geodesics, a generalization of the straight lines along which light travels in classical physics, of a certain space-time geometry. Given that, according to this theory, the curvature of spacetime is understood as the result of the presence of matter and radiation, this means that the light rays are deviated when approaching to any place with high matter concentration. The deflection of light in presence of massive bodies, and the phenomena resulting therefrom, are what we denominate *gravitational lensing*. This phenomenon increases the apparent brightness and angular size of the lensed sources, making it easier to study objects that would be otherwise too faint to be probed. Gravitational lensing effect, hereafter also denoted as lensing, can be occasioned by a number of massive objects, among which galaxy clusters, galaxies, and stars can be mentioned. These objects are called lenses or deflectors when they behave in such a way. Depending on the characteristics of these lenses, such as their mass and size, the effects of lensing on the observations of distant astronomical objects or faraway radiations will be different. Lensing effects can be observed at practically every angular length scale, and so three different lensing regimes are historically distinguished.

The lensing phenomena produced by extended lenses, like galaxies and galaxy clusters,

fall into two broad categories: strong lensing (SL) and weak lensing (WL). Strong lensing events may be observed if the observer, lens, and source are sufficiently well aligned along the line of sight. Since the late 30s it was already pointed out that galaxies could split images of background sources by a large enough angle so as to be observed (Zwicky, 1937). In such events, the light from the background source is deflected resulting in multiple images, highly distorted images, or more usually both things at the same time. For point-like background sources, there will be multiple images while for extended background emissions there often are distorted images, which can acquire the shape of rings or arcs. The extension of the source entails that the light from the background source is deflected differently, and hence the images produced are distorted. These distorsions can be in the radial or in the tangential direction. The most spectacular example of distorted images are the gravitational arcs observed near the center of many massive galaxy clusters (see Figures 0.2 and 3.1). All images belonging to the same background source carry the same spectral information, and thus multiple images can be identified thanks to an spectral analysis. In fact, the first gravitational lens system ever detected, discovered by Walsh et al. (1979), was confirmed due to the spectral similarity of the observed images. It consisted of two images of the quasar QSO 0957+561, separated by 6", seen through a foreground galaxy. The arrangement of these multiple images is determined by the mass distribution of the lens, and the relative position of the source. The most impressive example of multiple images is known as Einstein ring. This happens when the source, lens, and observer are all perfectly aligned, and the lens has circular symmetry, so the light is diverted forming a ring.

There is no accurate definition of weak lensing beyond the fact that the lenses have to be extended and that the angular separation between source and lens is large. Weak gravitational lensing maps source galaxies to new positions on the sky, systematically distorting their images. The effect of weak lensing can only be measured statistically because it shows up only by averaging on an ensamble of extended sources, which lay behind the lens. It is impossible to measure the effect on the images of single galaxies because the distorsion is indistiguishable from the intrinsic shape of the galaxies. A detailed review of weak lensing can be found in Bartelmann and Schneider (2001).

On the other hand, we have microlensing events, which are produced by lightweight masses whose size is small compared to its distance from us and to the source. Such lenses can be for example stars, planets or any other compact object floating in our or other galaxy. A mass able to bring about microlensing events is typically in the range $10^{-6} \leq M/M_{\odot} \leq 10^{6}$. The first observational detection of a microlensing event took place in 1989 (Vanderriest et al., 1989; Irwin et al., 1989). The presence of individual stars in the galaxies lensing quasars QSO 0957+561 and QSO 2237+0305 altered the magnification of

one of the quasar images with respect to the other. The occurrence of these events did not attract much attention until the first example of gravitational lensing was discovered. The fact is that it was not believed that microlensing could be observed owing to that the expected image separation was below the limit resolution. However, even when it is not possible to see the multiple images, the magnification may still be detected if the source and the lens move relative to each other. After the discovery of the first gravitational lens system, lensing by stellar mass objects become more appealing. Paczynski (1986) contributed to this rise by showing that in any nearby galaxy at any time one star out of a million might be measurably magnified by an intervening star in the halo of our galaxy. Initially, it was believed that the individual monitoring of the light curves of a million stars would never be feasible. Nevertheless, technological advances have made it possible to detect nowadays the order of $\sim 10^4$ microlensing events, and have allowed the study of this phenomenon to become a research field itself. It is worth to mention the discovery, through microlensing, of the farthest star ever detected, called Icarus, at $z = 1.49$ (Kelly et al., 2018). A detailed review of microlensing can be found in Wambsganss (2006).

Throughout the last decades, gravitational lensing has proven to be a major tool for observational cosmology, and has given rise to an independent field of research within the cosmology community. Microlensing techniques are used to make precise measurements of the masses of lens stars that do not rely on any assumption about their internal physics. These measurements also help in the validation of theoretical stellar models. One of the theories proposed for dark matter is that it could exist in the form of compact objects distributed in the halo of the galaxy. If these objects, known as MACHOs, have a suitable mass, they could originate detectable microlensing signals from background stars in nearby galaxies (Paczynski, 1986). Several searches for dark matter are ongoing following this idea. Microlensing events are also used for the search of exoplanets (Tsapras, 2018). The gravitational influence of a transiting planet on the light coming from its host star can be detected in the form of brightness variations. The first microlensing planet was discovered in 2004 (OGLE Collaboration, 2004), and ever since ~ 50 exoplanets have been found through this technique, including cool low mass planets (Beaulieu et al., 2006). Microlensing has developed into an exciting method for studying the nature and distribution of mass in our galaxy.

Focusing now on strong lensing, we can name two relatively distinct fields: lensing by galaxy clusters and lensing by galaxies. These events are different both in appearance and in abundance. Galaxy-cluster systems are far less common than their galaxy lens counterparts. The baryons in galaxies have had enough time to cool and shape the visible galaxy, thereby giving a cuspy profile, suitable for lensing. The density profiles of galaxy clusters are not usually so cuspy, making them less ideal lenses. Therefore, only the most

dense clusters have surface densities large enough for lensing. For this reason, despite not being as numerous as galaxy-galaxy systems, they are much more impressive. While galaxy lens systems have typical image separations of a few arcseconds, the most massive clusters are able to produce multiple images of distant galaxies with separations of up to $\sim$ 1 arcmin. The study of these images and its distorsion can provide us with unique information about the distribution of baryonic and dark matter within the foreground mass acting as a lens. The more complicated is the gravitational potential of the lens, the harder is to extract the information. This information allows us, for instance, to constrain galaxy mass models, the stellar initial mass function or the abundance of dark matter subhalos. The most accurate estimates of the matter distribution in the central region of galaxies are, indeed, due to gravitational lensing analyses. The magnification effect derived from gravitational lensing works like a cosmic telescope, and has enabled us to observe distant objects which would otherwise remain undetected. The statistics of lensing events also turn them into a promising way of inferring or deriving significant constraints on the cosmological parameters (Kochanek, 1993, 1996; Maoz and Rix, 1993; Bartelmann et al., 1998; Falco et al., 1998; Link and Pierce, 1998; Golse and Kneib, 2002; Chiba and Futamase, 1999). The cosmological parameter best determined through lensing is the Hubble constant. Refsdal (1964) was the first one who pointed out that the Hubble constant could be estimated by measuring the time delay between the appearance of the images of an intrinsically variable lensed source, such as SNe. Since then, H_0 has been measured through the analysis of many lensed sources (Kundić et al., 1997; Schechter et al., 1997; Biggs et al., 1999; Koopmans et al., 2003; Kochanek and Schechter, 2004), but it was not until recently that SNe were finally used for this purpose (Vega-Ferrero et al., 2018). These analyses have improved the precision of other probes (Wong et al., 2020) due, for example, to the fact that this approach works at larger redshifts than other methods.

Weak gravitational lensing has been widely used as a direct probe of the matter distribution in our Universe on different scales (Zhang, 2011), including large-scale structure, galaxy clusters, and galaxies (Wilson et al., 1996; Umetsu, 2011). Looking at large scales, we found that the analysis of weak lensing by large-scale structures can provide direct measurements of the primordial power spectrum of the density fluctuations in the Universe (Hilbert et al., 2012). The CMB is affected by gravitational lensing as well: it changes the CMB power spectra, induces non-Gaussianities, and generates a B-mode polarization signal. The study of all these effects is a well-developed probe of large-scale clustering over a broad range of redshifts (Lewis and Challinor, 2006; Challinor et al., 2018). Reaching at last the largest possible scale, we found that the entire Universe acts as a lens. Its energy density distorts the shapes of any astronomical object and any radiation all over the sky. The signal produced by these effects, known as cosmic-shear, can be detected, if treated

Figure 3.1: Giant arc observed by Hubble Space Telescope in an image of one of the most massive galaxy clusters known, Abell 1689. Courtesy of NASA, N. Benitez (JHU), T. Broadhurst (The Hebrew University), H. Ford (JHU), M. Clampin (STScI), G. Hartig (STScI), G. Illingworth (UCO/Lick Observatory), the ACS Science Team and ESA.

statistically, in large field-of-view observations. This signal can be exploited to constrain the structure growth of the Universe in a very powerful and direct way.

3.2 Theory of lensing

3.2.1 Light deflection

Although it can be observed on different length scales, the basic explanation behind lensing is always the same and can be approached in a simple way. In this work we only discuss the effects of lensing at first order. A description of the second order effects on the images of lensed sources can be read in Meneghetti (2016). Assuming at first a point mass lens, whose Schwarzschild radius is:

$$R_S = \frac{2GM}{c^2}, \tag{3.1}$$

the deflection angle $\hat{\alpha}$ of a light ray crossing its gravitational field is given by:

$$\hat{\alpha}(\xi) = \frac{4GM}{c^2\xi} = \frac{2R_s}{\xi} \tag{3.2}$$

where M is the mass of the point mass lens, G is the universal gravitational constant, c is the speed of light, and ξ is the impact parameter, i.e., the distance from the deflector at which an unperturbed light ray would pass. This expression is true, provided that the impact parameter is much larger than the Schwarzschild radius of the lens. As can be seen, $\hat{\alpha}$ does not depend on the wavelength of the light. Therefore, gravitational lensing is inherently achromatic, and any chromatic effect is only consequence of the wavelength-dependent geometry of the different emission regions in the Universe.

The deflection angle shown in Eq. (3.2) depends linearly on the mass of the lens, thus the deflection angles of a set of lenses can be linearly superposed. Supposing we have a scattered distribution of N point masses on the same plane, the deflection angle of a light ray propagating through the plane at $\boldsymbol{\xi}$ is:

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \sum_i \hat{\boldsymbol{\alpha}}_i(\boldsymbol{\xi} - \boldsymbol{\xi}_i) = \frac{4G}{c^2} \sum_i M_i \frac{\boldsymbol{\xi} - \boldsymbol{\xi}_i}{|\boldsymbol{\xi} - \boldsymbol{\xi}_i|^2} \tag{3.3}$$

with $\boldsymbol{\xi}_i$ and M_i being the positions in the plane and masses, respectively, of the $1 \leq i \leq N$ point sources.

As we know, distributions of matter in the Universe cannot be properly modeled as point masses, so more realistic lens models need to be considered. Despite being such massive structures, galaxies and galaxy clusters are concentrated on a relatively small volume, compared to the distance that light from a distant source has to travel to reach us. When the distances between observer, lens and source are much larger than the physical size of the lens, we can assume that the entire matter distribution of the lens is contained in one plane, called the *lens plane*. The sources are also assumed to lie on a plane, called the *source plane*. This is known as the *thin screen approximation*, and is sketched in Fig. 3.2. The lensing matter distribution under this approximation is fully described by its surface mass density:

$$\Sigma(\boldsymbol{\xi}) = \int \rho(\boldsymbol{\xi}, z) dz \tag{3.4}$$

where ρ is the three-dimensional density mass of the lens, $\boldsymbol{\xi}$ is now defined as a two-dimensional vector on the lens plane, and z indicates the direction along which unperturbed light rays propagate. Under the thin screen approximation, light deflection occurs on a short section of the light path, and is described by the following angle:

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \int \frac{(\boldsymbol{\xi} - \boldsymbol{\xi'})\Sigma(\boldsymbol{\xi'})}{|\boldsymbol{\xi} - \boldsymbol{\xi'}|^2} d^2\xi' \tag{3.5}$$

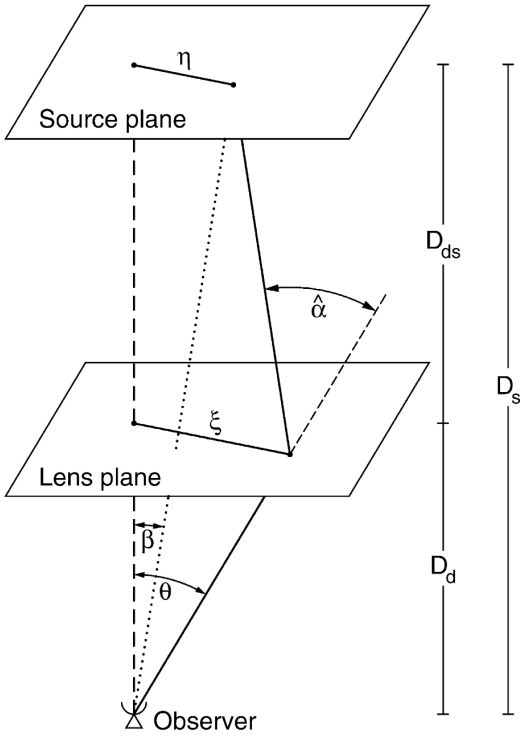


Figure 3.2: Sketch of a gravitational lens system, from Bartelmann and Schneider (2001).

3.2.2 Lens equation

The geometry of a typical gravitational lens system, assuming the thin screen approximation, is shown in Fig. 3.2. The optical axis, depicted as a dashed line perpendicular to the source and lens planes and centered in the observer, will be used henceforward as reference. Let suppose a source placed in the source plane at transverse distance $\boldsymbol{\eta}$ from the optical axis. A light ray coming from that source intersects the lens, in the lens plane, at transverse distance $\boldsymbol{\xi}$ from the optical axis. This light ray is deflected by an angle $\hat{\boldsymbol{\alpha}}$ defined in Eq. (3.5), so the observer receives the light coming from the source as if it was emitted at a different position. The angular separation between the optical axis and the true source position is $\boldsymbol{\beta}$, while the angle between the optical axis and the image of the source seen by the observer is $\boldsymbol{\theta}$. The angular diameter distances from the observer to the lens, to the source, and between the lens and the source are D_L, D_S, and D_{LS}, respectively.

Knowing this, we can express the transverse distances $\boldsymbol{\eta} = \boldsymbol{\beta} D_S$ and $\boldsymbol{\xi} = \boldsymbol{\theta} D_L$ in terms of their angular positions. As long as $\hat{\boldsymbol{\alpha}}$, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are small, the true source position and its observed position are related by the following geometrical expression:

$$\boldsymbol{\beta} D_S = \boldsymbol{\theta} D_S - \hat{\boldsymbol{\alpha}} D_{LS} \tag{3.6}$$

It is now convenient to define the reduced (or scaled) deflection angle:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) \equiv \frac{D_{LS}}{D_S} \hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}) \tag{3.7}$$

so Eq. (3.6) can be expressed as

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}) \tag{3.8}$$

This geometrical relation, called the *lens equation*, is the centerpiece of gravitational lensing, that allows us to describe the ray-tracing between the source plane and the lens plane. It is non-linear since multiple images $\boldsymbol{\theta}$ can result from a single source position $\boldsymbol{\beta}$. This equation holds, provided that the Euclidean relation *physical size* = *angular size* × *distance* is fulfilled. Although it is not guaranteed that this relation meets whether the spacetime is curved, it is possible to define distances in a curved spacetime such that satisfy this relation.

3.2.3 Lensing potential

Before proceeding further, it is useful to define a dimensionless quantity equivalent to the surface mass density, known as the *convergence* κ:

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(\boldsymbol{\theta} D_L)}{\Sigma_{cr}} \tag{3.9}$$

with $\Sigma(\boldsymbol{\theta} D_L)$ being the surface mass density from Eq. (3.4), and Σ_{cr} is the critical surface density defined as

$$\Sigma_{cr} = \frac{c^2}{4\pi G} \frac{D_S}{D_L D_{LS}} \tag{3.10}$$

For realistic cases, such as the cluster analyzed in the next chapter, Σ_{cr} can take values of

the order of ≈ 0.4 g/cm^2. This means that in the center of a massive cluster, the projected mass is comparable to that of a 1€ coin or 50 sheets of stacked paper.

The strong lensing effect often requires that $\Sigma(\boldsymbol{\xi}) > \Sigma_{cr}$. Recalling Eq. (3.5), the reduced deflection angle under the thin screen approximation now can be read as

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi} \int \kappa(\boldsymbol{\theta'}) \frac{\boldsymbol{\theta} - \boldsymbol{\theta'}}{|\boldsymbol{\theta} - \boldsymbol{\theta'}|^2} d^2\theta' \tag{3.11}$$

The efective lensing potential of an extended mass by applying the thin screen approximation can be written is terms of the convergence as

$$\psi(\boldsymbol{\theta}) = \frac{1}{\pi} \int \kappa(\boldsymbol{\theta'}) \ln|\boldsymbol{\theta} - \boldsymbol{\theta'}| d^2\theta' \tag{3.12}$$

And the gradient of this potential allows us to obtain the reduced deflection angle

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \boldsymbol{\nabla}\psi(\boldsymbol{\theta}) \tag{3.13}$$

while its Laplacian gives us the aforementioned convergence:

$$\kappa(\boldsymbol{\theta}) = \frac{1}{2}\Delta\psi(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\nabla}^2\psi(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\nabla}\boldsymbol{\alpha}(\boldsymbol{\theta}) \tag{3.14}$$

3.2.4 Distorsion and magnification

One of the main effects of gravitational lensing is that it warps the shape of the sources observed. This can be negligible for unresolved sources but not for extended sources. The light coming from different points of an extensive source is deflected differently resulting in distorted images. The most spectacular example of distorsion are the arcs, belonging to background galaxies, observed near the center of massive galaxy clusters. As we saw in Sec 3.2.2, the relation between $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ comes determined by the lens equation, so the true shape of the sources can be derived by solving the lens equation for all the points within its images. On condition that the size of the source is insignificant compared to the scales on which the properties of the lens change, which is virtually always fulfilled, the lens equation can be locally linearized. The distorsion of the images is, therefore, described at first order by the following Jacobian matrix:

$$\mathcal{A}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \left(\delta_{ij} - \frac{\partial^2 \psi(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right) \tag{3.15}$$

where θ_i and θ_j indicate the two components of $\boldsymbol{\theta}$ on the lens plane, and δ_{ij} is the Kronecker delta. For the sake of clarity we will use the following notation:

$$\psi_{ij} \equiv \frac{\partial^2 \psi(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \tag{3.16}$$

If we retrieve Eq. (3.14), we can rewrite the convergence in terms of its partial derivatives with this new notation:

$$\kappa(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\nabla}^2 \psi(\boldsymbol{\theta}) = \frac{1}{2}\Big(\frac{\partial^2}{\partial \theta_1^2} + \frac{\partial^2}{\partial \theta_2^2}\Big)\psi = \frac{1}{2}(\psi_{11} + \psi_{22}) \tag{3.17}$$

We can now introduce another important magnitude in gravitational lensing, known as the *shear*:

$$\gamma = (\gamma_1^2 + \gamma_2^2)^{1/2} = \gamma_1 + i\gamma_2 = |\gamma| e^{2i\phi} \tag{3.18}$$

The components of this pseudo-vector, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$, like the convergence, are defined as a function of the partial derivatives of the lensing potential:

$$\gamma_1 = \frac{1}{2}\Big(\frac{\partial^2}{\partial \theta_1^2} - \frac{\partial^2}{\partial \theta_2^2}\Big)\psi = \frac{1}{2}(\psi_{11} - \psi_{22}) \equiv \gamma(\boldsymbol{\theta}) \cos 2\phi(\boldsymbol{\theta}) \tag{3.19}$$

$$\gamma_2 = \frac{\partial^2}{\partial \theta_1 \theta_2}\psi = \psi_{12} = \psi_{21} \equiv \gamma(\boldsymbol{\theta}) \sin 2\phi(\boldsymbol{\theta}) \tag{3.20}$$

With the previous definitions, the Jacobian matrix from Eq. (3.15) can be written as follows:

$$\begin{aligned} \mathcal{A}(\boldsymbol{\theta}) &= \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_2 \end{pmatrix} \\ &= (1 - \kappa)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \gamma \begin{pmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{pmatrix} \end{aligned} \tag{3.21}$$

The physical meaning of the convergence and the shear is much clearer once this form of the matrix $\mathcal{A}$ is reached. The convergence (κ) induces isotropic distortions in the images,

i.e., the source is mapped onto an image with the same shape but its size is rescaled by a constant factor in all directions. On the other hand, the shear (γ) induces an anisotropic distorsion, i.e., the intrinsic shape of the source is stretched along one preferred direction. The quantity γ, defined in Eq. (3.18), determines the magnitude of the shear while its orientation is described by the angle ϕ. For this reason, a circular source like the one shown in Fig. 3.3 is mapped onto an elliptical image after being affected by κ and γ. The ellipticity of this image can be defined as

$$\epsilon = \epsilon_1 + i\epsilon_2 = \frac{1-(b/a)}{1+(b/a)} e^{2i\phi} \tag{3.22}$$

where ϕ is the orientation angle of the image, a and b are its major and minor axes:

$$a = \frac{r}{1-\kappa-\gamma}, \quad b = \frac{r}{1-\kappa+\gamma}, \tag{3.23}$$

and r is the radius of the circular source. Therefore, we can see how the observed ellipticities allow measuring a combination of κ and γ:

$$\langle \epsilon \rangle = \left\langle \frac{\gamma}{1-\kappa} \right\rangle \tag{3.24}$$

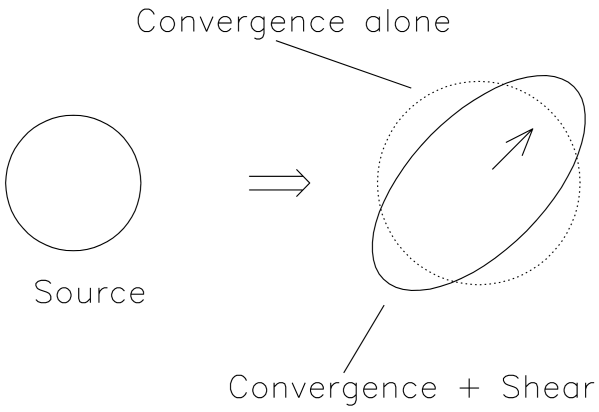


Figure 3.3: Distorsion effects on a circular source due to the convergence and shear. Taken from Narayan and Bartelmann (1996).

Image distortion is inherently associated with a change in the flux that we receive from the source. The surface brightness of the source is preserved in gravitational lensing, since no photons are absorbed or emitted. Nevertheless, the change in the solid angle

under which we observe the source results in an increase, or decrease, of the flux coming from the source. The ratio between the fluxes of the observed image and the unlensed source is known as the *magnification*. The magnification is calculated as the inverse of the determinant of the Jacobian matrix $\mathcal{A}$ from Eq. (3.21):

$$\mu = \det\mathcal{M} = \frac{1}{\det\mathcal{A}} = \frac{1}{(1-\kappa)^2 - |\gamma|^2} \tag{3.25}$$

where $\mathcal{M} = \mathcal{A}^{-1}$ is the magnification tensor. The eigenvalues of $\mathcal{M}$ are given by

$$\mu_r = \frac{1}{\lambda_r} = \frac{1}{1-\kappa+\gamma} \tag{3.26}$$

$$\mu_t = \frac{1}{\lambda_t} = \frac{1}{1-\kappa-\gamma} \tag{3.27}$$

and allow measuring the magnification in the radial direction (μ_r) and in the tangential direction (μ_t).

The magnification is ideally infinite when $\det\mathcal{A} = 0$. The set of points in the lens plane ($\boldsymbol{\theta}$-space) where this condition is met form closed curves called *critical lines*, and the corresponding positions in the source plane ($\boldsymbol{\beta}$-space) are called *caustics*. Therefore, there are two critical curves: the *radial* critical line, where $\mu_r = 0$, and the *tangential* critical line, for which $\mu_t = 0$. They are linked, respectively, to a radial caustic and a tangential caustic. A source placed near a radial caustic will produce images close to the radial critical line, and stretched perpendicular to it. And any circularly symmetric source placed around the tangential caustic will form strongly distorted images along the tangential critical line. In addition to highlight regions of high magnification, critical lines and caustics also separate regions of different image multiplicities. Sources placed within a caustic are multiply imaged, and the number of images changes as the source moves across the caustic network. A lens capable of inducing such effects in the source images is called critical or supercritical. Both image distortions and the image multiplicity are features that belong to the strong lensing regime. Strong lensing often occurs in the central regions of galaxies and galaxy clusters. Galaxy clusters with high mass concentration occasionally produce giant gravitational arcs when a background galaxy is aligned with one of the cluster caustics (see Figures 0.2 and 3.1).

3.3 Lens models

It is a common practice to use analytical lens models to reproduce the behavior of astronomical objects as lenses. They are easily understandable and the derivation of their lensing properties is pretty straighforward. Compact objects like planets, stars or MACHOs are usually well aproximated by point mass lenses. The next rung of the ladder are the axially symmetric (or circularly symmetric) lens models, for which a good choice of the optical axis allows reducing the lensing analysis to a one-dimensional problem. However, they are unable to describe well enough the lens behavior of most astrophysical sources. The introduction of ellipticity in such models usually solves that problem. Elliptical lens models are the best choice for approximating elliptical galaxies, which are a representative example of lenses in the Universe (Kochanek, 1993). In this section we will approach some of these lens models.

3.3.1 Point mass lens

We already introduced point mass lenses at the beginning of the previous section. A point mass is the simplest example of an axially symmetric lens model. Recalling the expression of $\hat{\boldsymbol{\alpha}}$ from Eq. (3.2), and after applying Eq. (3.7), we can derive the reduced deflection angle of a point mass as a function of the angular distance $\boldsymbol{\theta}$:

$$\alpha(\boldsymbol{\theta}) = \frac{4GM}{c^2\boldsymbol{\theta}}\frac{D_{LS}}{D_L D_S} \tag{3.28}$$

The expression of its lensing potential can be easily obtained by using Eq. (3.13):

$$\psi(\boldsymbol{\theta}) = \frac{4GM}{c^2}\frac{D_{LS}}{D_L D_S}\ln|\boldsymbol{\theta}| \tag{3.29}$$

The lens equation, Eq. (3.8), of a point source lensed through this point mass lens can now be read as

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \frac{4GM(\boldsymbol{\theta})}{c^2|\boldsymbol{\theta}|}\frac{D_{LS}}{D_L D_S} \tag{3.30}$$

Owing to the rotational symmetry of a point mass lens, a source located on the optical axis ($\boldsymbol{\beta} = 0$) is imaged as a ring, provided that the lens is supercritical. This symmetry property is shared by the rest of circularly symmetric lens models we will see. The radius

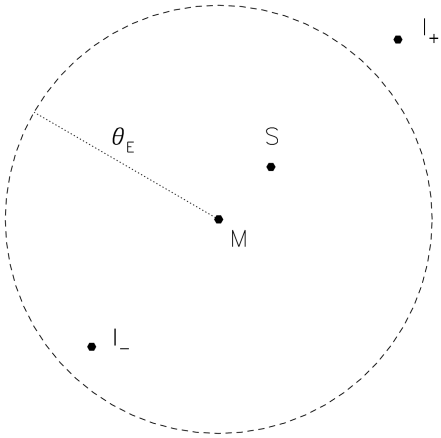


Figure 3.4: Locations of the source (S) and its images (I_+ and I_-) after being lensed by a point mass M. Given the position of the source, one image falls inside the Einstein ring and the other outside, both on opposite sides of the source. The Einstein ring is depicted as a dashed circle of radius θ_E. Taken from Narayan and Bartelmann (1996).

of such ring for a point mass is obtained by setting $\boldsymbol{\beta} = 0$ in the above equation:

$$\theta_E \equiv \sqrt{\frac{4GM}{c^2}\frac{D_{LS}}{D_L D_S}} = \sqrt{\frac{4GM}{c^2 D}} \tag{3.31}$$

where $D \equiv \frac{D_L D_S}{D_{LS}}$ is defined as the effective lensing distance. Eq. (3.31) is known as the *Einstein radius*, and depends on both the properties of the lens and the distances involved. A point mass lens causes two images for any source, regardless of its angular distance $\boldsymbol{\beta}$ to the lens. The typical angular separation between these two images is roughly $2\theta_E$. The angular positions of these two images are given by

$$\theta_\pm = \frac{1}{2}\left(\beta \pm \sqrt{\beta^2 + 4\theta_E^2}\right) \tag{3.32}$$

The locations of these two images for a given source position $\boldsymbol{\beta} \neq 0$, using the Einstein radius as a reference, are shown in Fig. 3.4. Each image is located on one side of the source, being one inside the Einstein radius and the other outside. As the source moves aways from the lens, one of the images approaches to the true position of the source tending to a value-one magnification, while the other image approaches to the lens, becoming fainter and fainter. Sources located far outside the Einstein ring are slightly magnified, whereas sources at a distance from the optical axis lower than θ_E are heavily magnified.

3.3.2 Axially symmetric lenses

The great advantage of using axially symmetric lenses, of which the point mass lens model is also part, is that their surface mass density does not depend on the position angle $\boldsymbol{\theta}$ to the lens center. It only depends on its modulus $|\boldsymbol{\theta}|$. Therefore, the optical axis can be shifted to the center of symmetry of the lens and the light deflection reduces to a one-dimensional problem. Following the same procedure as for the point mass, we can derive the expressions of the reduced deflection angle and the lensing potential for a point source being lensed by a circularly symmetric lens with an arbitrary mass profile:

$$\alpha(\boldsymbol{\theta}) = \frac{4GM(\boldsymbol{\theta})}{c^2|\boldsymbol{\theta}|}\frac{D_{LS}}{D_L D_S}, \quad \psi(\boldsymbol{\theta}) = \frac{4GM(\boldsymbol{\theta})}{c^2}\frac{D_{LS}}{D_L D_S}\ln|\boldsymbol{\theta}| \tag{3.33}$$

The Einstein radius corresponding to this lens model is again obtained by setting $\boldsymbol{\beta} = 0$ in the lens equation:

$$\theta_E \equiv \sqrt{\frac{4GM(\theta_E)}{c^2 D}} \tag{3.34}$$

As discussed above, multiple images can be produced if the lens is critical. The number of images depends on the position of the source with respect to the caustics, and the Einstein radius defines the scale to measure the angular separation between them. In axially symmetric lenses, the tangential critical line is mapped onto a single point caustic, whereas the radial critical curve, which corresponds to the Einstein ring, is mapped onto a closed circular radial caustic on the source plane. As it can be seen in Fig. 3.5, sources located inside the radial caustic produce three images, while sources that lie outside the radial caustic only produce one image. As the tangential caustic consists of a single point at $\boldsymbol{\beta} = 0$, the tangential critical line do not affect to the number of images observed. It is proven that the number of images produced in axially symmetric lenses is always odd. The only exception are the point lenses, i.e., lenses with infinity density at the center. As we show in Eq. (3.32) and Fig. 3.4, only two images can be observed when the source lie within the radial caustic. A detailed description of the image multiplicity in axially symmetric lenses can be found in Schneider et al. (1992). The distortion effects in the images produced by this type of models are better noted when they perform on extended sources. We will discuss this in more detail after bringing in ellipticity to these models in Sec. 3.3.4.

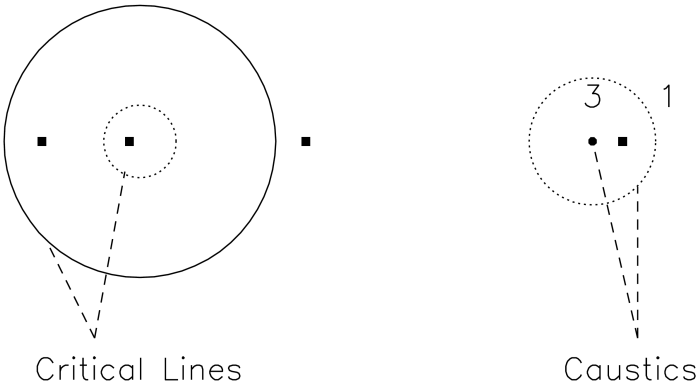


Figure 3.5: Imaging of a point source lensed by a non-singular, circularly symmetric lens. *Right*: Positions of the source (black square) and the lens (black circle), along with the radial caustic depicted using a dashed line. The tangential caustic consists of a single point, where the source is placed. *Left*: Image positions of the source (three images because the source is within the Einstein ring) along with the critical lines. The outer curve is the tangential critical line while the inner curve is the radial critical curve. Taken from Narayan and Bartelmann (1996).

3.3.2.1 Singular Isothermal Sphere

If we assume that the matter content within the lens behaves as an ideal gas, bound by its own spherically symmetric gravitational potential, in thermal and hydrostatic equilibrium, we can derive the following density distribution

$$\rho(r) = \frac{\sigma_v^2}{2\pi G}\frac{1}{r^2} \tag{3.35}$$

where r is the distance from the lens center. This density profile is known as the *singular isothermal sphere*, and enables to relate the underlying mass distribution of the lens to the velocity dispersion σ_v of its individual gas components. The derivation of this density distribution can be seen in Narayan and Bartelmann (1996) and Binney and Tremaine (2008). By projecting this density along the line of sight, we obtain the corresponding surface mass density in terms of the impact parameter ξ:

$$\Sigma(\xi) = \frac{\sigma_v^2}{2G}\frac{1}{\xi} \tag{3.36}$$

The Einstein radius of the singular isothermal sphere is given by

$$\theta_E = 4\pi \frac{\sigma_v^2}{c^2} \frac{D_{LS}}{D_S} \tag{3.37}$$

The reduced deflection angle, lensing potential, convergence, shear, and magnification for this lens model can now be easily written as functions of the Einstein radius:

$$|\boldsymbol{\alpha}| = \theta_E, \quad \psi = \theta_E |\boldsymbol{\theta}|, \quad \kappa = \gamma = \frac{\theta_E}{2|\boldsymbol{\theta}|}, \quad \mu = \frac{|\boldsymbol{\theta}|}{|\boldsymbol{\theta}| - \theta_E} \tag{3.38}$$

Multiple images are obtained only if the source lies inside the Einstein ring ($\boldsymbol{\beta} < \theta_E$). Under this circumstance, the singular isothermal sphere model generates only two images than can be detected. A third image is theoretically predicted, but it is located at $\boldsymbol{\theta} = 0$ and has zero flux. As already mentioned above, the angular separation between the two images is $2\theta_E$. If the source lies outside the Einstein ring ($\boldsymbol{\beta} < \theta_E$), only one image is produced.

3.3.2.2 Softened Isothermal Sphere

Even though the singular isothermal sphere reproduces in a natural way the flat rotation curves of galaxies, its density distribution is singular at $r = 0$ and tends to infinity for $r \to \infty$. For this reason, to avoid this problem, we can either look for more elaborate models or replace the singularity at the center of the lens by a core region with finite density. If we decide to follow this last path, we have a model called *softened isothermal sphere* (SIS). The deflection angle, lensing potential, convergence, and shear components of this lens model are as follows:

$$\alpha(\boldsymbol{\theta}) = \theta_E \frac{\boldsymbol{\theta}}{(\boldsymbol{\theta}^2 + \theta_c^2)^{1/2}} \tag{3.39}$$

$$\psi(\boldsymbol{\theta}) = \theta_E \, (\boldsymbol{\theta}^2 + \theta_c^2)^{1/2} \tag{3.40}$$

$$\kappa(\boldsymbol{\theta}) = \theta_E \frac{\boldsymbol{\theta}^2 + 2\theta_c^2}{2(\boldsymbol{\theta}^2 + \theta_c^2)^{3/2}} \tag{3.41}$$

$$\gamma_1(\boldsymbol{\theta}) = -\theta_E \frac{\boldsymbol{\theta}^2}{2(\boldsymbol{\theta}^2 + \theta_c^2)^{3/2}} \cos 2\phi \tag{3.42}$$

$$\gamma_2(\boldsymbol{\theta}) = -\theta_E \frac{\boldsymbol{\theta}^2}{2(\boldsymbol{\theta}^2 + \theta_c^2)^{3/2}} \sin 2\phi \tag{3.43}$$

where θ_c is the core radius and θ_E is the Einstein radius from Eq. (3.37). Now that the central singularity has vanished, the third image produced when the source is inside the Einstein ring can be observed.

3.3.3 The Navarro-Frenk-White density profile

Navarro, Frenk, and White found, thanks to numerical simulations of the formation and evolution of dark matter halos, that its density distribution can be very well fitted by a certain profile. The Navarro-Frenk-White (NFW) density profile (Navarro et al., 1996, 1997) is since then the most commonly used profile for modeling the mass distribution of dark matter in galaxy clusters (Bartelmann, 1996; Wright and Brainerd, 2000; Golse and Kneib, 2002; Meneghetti et al., 2003). The density of dark matter as a function of the radius in the NFW profile is given by

$$\rho(r) = \frac{\rho}{(r/r_s)(1 + r/r_s)^2} \tag{3.44}$$

where the two parameters, r_s and ρ_s, are the scale radius and the characteristic density of the halo, respectively. The logarithmic slope of this density profile changes from -1 at the center (flatter in the inner region) to -3 at large radii. It is flatter than the SIS density profile in the inner part of the halo, and steeper in the outer regions. NFW models parametrize dark matter halos by their masses M_{200}, i.e., the mass within a sphere of radius r_{200}, in which the average density is 200 times the critical density. The concentration parameter of a dark matter halo can be defined then as follows:

$$c_{200} = \frac{r_{200}}{r_s} \tag{3.45}$$

There is an extension of the NFW density profile, known as the generalized NFW (gNFW) model (see Zhao 1996; Jing and Suto 2000; Wyithe et al. 2001), which is defined as follows

$$\rho_{gNFW}(r) = \frac{\rho_s}{(r/r_s)^{\gamma_{gNFW}}(1 + r/r_s)^{3-\gamma_{gNFW}}} \tag{3.46}$$

As can be seen, if $\gamma_{gNFW} = 1$ the original NFW model is recovered, and $\gamma_{gNFW} < 1$ or

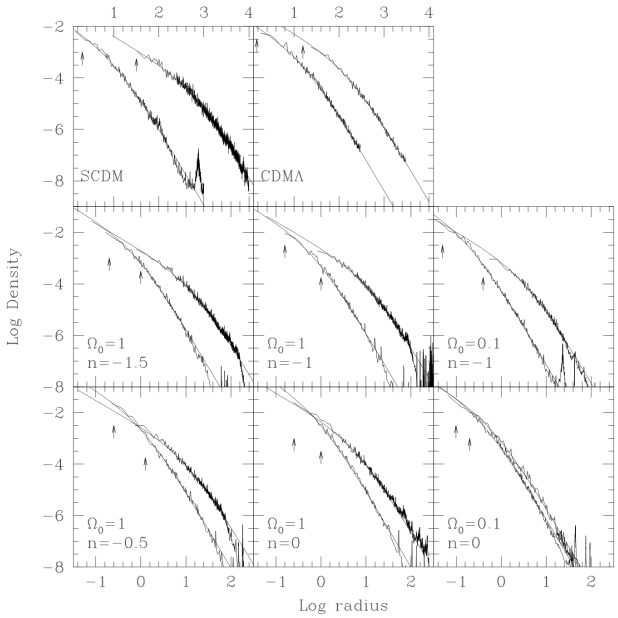


Figure 3.6: Density profiles of two simulated dark matter halos from Navarro et al. (1996), fitted each one to a NFW profile.

$\gamma_{gNFW} > 1$ correspond, respectively, to shallower or steeper density profiles.

3.3.4 More realistic lens models

In order to achieve lens models capable of describing the properties of real lenses more realistically, Blandford and Kochanek (1987) recommended to introduce elliptical lensing potentials. This can be done in a simple way by performing the following substitution on the models described above:

$$|\boldsymbol{\theta}| \rightarrow \sqrt{\frac{\theta_1^2}{1-\epsilon} + (1-\epsilon)\theta_2^2} \tag{3.47}$$

where ϵ = 1 - b/a is the ellipticity, and a and b are the major and minor axis of the ellipse. The resulting model will have elliptical iso-density contours whose major axis will be oriented along the θ_2-direction, but can as well be rotated any desired angle. The tangential and radial caustics were a point and a circle, respectively, for the axially symmetric models, whereas the critical lines consisted of circles on the lens plane. Increasing the ellipticity of the lens strengthens its shear, and the caustics and the critical lines expand their areas and change accordingly. The caustics tend to acquire the shape of cusps, while the critical

lines assume a dumbbell shape. This last is due to the fact that such is the characteristic shape of the surface mass density in elliptical lenses.

The growth of the caustic network entails an increase in the variaty of image configurations that might be produced. Some examples of the effect of elliptical lens models on compact and extended sources are shown in Fig. 3.7. Each row in that figure corresponds to the same lens and source configuration. The caustics in the source plane are shown on the right, and the images produced together with the critical lines are shown on the left. A source enclosed by both caustics, and placed behind the lens center, generates five images: four bright images forming a cross-shaped pattern, and another one fainter appears at the lens center. When a compact source moves away from the lens center along a symmetry line (top right panel), two outer images begin to approach to each other, then merge, and finally dissapear (top left panel) when the source crosses the tangential caustic. A third outer image and the central image start to approach to each other as well, and dissapear when the source crosses the radial caustic. During this process, the fourth exterior image is coming near the true source position while diminishing its distorsion. It is the last image left. On the other hand, while a source with a larger angular size moves towards a cusp edge (bottom right panel), three outer images merge forming a single elongated arc (bottom left panel). This happens just when the source is crossing the tangential caustic. This large arc-like image shortens and approaches to the position of source as the source keeps moving away from the center. In the end, this image will be the last one observed, and it will be placed virtually at same position than the source.

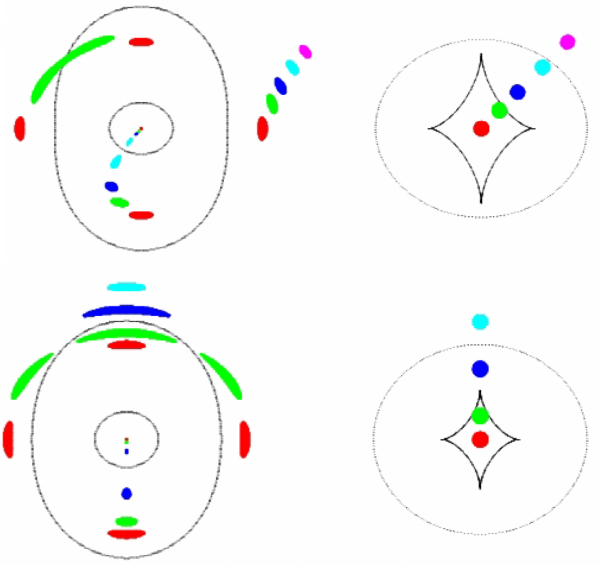


Figure 3.7: The upper panel shows the movement of a compact source towards a inter-cusp segment of a tangential caustic produced by an elliptical lens (right draw), and the resulting images depending on the position of the source (left draw). The lower panel shows the same, but the source is larger and is moving towards one of the cusps of the tangential caustic. Taken from Narayan and Bartelmann (1996).

A last aspect to consider is the gravitational influence of the matter in the vicinity of the lens on the light deflection. This can be taken into account by adding an external shear γ_e, whose lensing potential and deflection angle are given by

$$\psi_\gamma(\boldsymbol{\theta}) = \frac{\gamma_e}{2}(\theta_1^2 - \theta_2^2), \quad \boldsymbol{\alpha}(\boldsymbol{\theta}) = \gamma_e(\theta_1 - \theta_2) \tag{3.48}$$

Since the SIS density profile is significantly steeper in the central region than the NFW profile, this latter is much more sensitive to the incorporation of external perturbations.

3.4 Strong lensing in galaxy clusters

Strong gravitational lensing is a powerful technique for probing galaxy mass distributions and for measuring cosmological parameters. In this thesis we aim at estimating the mass distribution of a galaxy cluster through an analysis of its strong lensing observables. By and large, a mass reconstruction based on strong lensing constraints can be treated as a multipole expansion of the convergence:

$$\kappa(\theta) = \kappa_0(\theta) + \sum_{n=1}^{\infty} \kappa_n(\theta) e^{in\phi} \tag{3.49}$$

where depending on to which order we expand, the cluster structure will be described in more or less detail. The monopole term κ_0 describes an axially symmetric lens, while the quadrupole depicts the degree of ellipticity of the lens. In addition to this, one can add an external shear, as defined in Eq. (3.48), in order to account for perturbations coming from the outer regions of the cluster.

Once all the strong lensing observables have been gathered together, the mass reconstruction of the cluster can be tackled from two different approaches. The most broadly used one is the parametric approach, where a certain model is assumed in order to describe the behavior of the cluster as a lens (Kneib et al., 1993, 1995, 1996; Broadhurst et al., 1995; Natarajan and Kneib, 1997; Sand et al., 2002; Gavazzi et al., 2004; Halkola et al., 2006; Limousin et al., 2010). The model parameters p are related to all the unknowns in the lensing problem, such as the positions β of the background galaxies and the mass distribution of the cluster. The lens equation for each one of the N observations is as follows:

$$\beta_i = \theta_i - \alpha(\theta_i, p) \quad 1 \leq i \leq N \tag{3.50}$$

Parametric models are based on building, using the lens equation above, a χ^2 function in a similar formulation to the following:

$$\chi^2_\theta = \sum_{i=1}^{N} \left(\frac{\theta_i^{\rm obs}(\beta) - \theta_i^{\rm mod}}{\sigma_i} \right)^2 \tag{3.51}$$

The solution of the strong lensing problem is found by minimizing the corresponding χ^2_θ function with respect to the model parameters, i.e., looking for a model whose predicted images ($\theta_i^{\rm mod}$) are close enough to the observed ones ($\theta_i^{\rm obs}$). The results obtained will be as accurate as the chosen parametrization and the quality of the data allow it. Parametric models are the best choice for modeling strong lensing clusters when the number of lensing constraints is relatively small. Its wide use comes from the fact that, in the early years of strong lensing data analysis, it was common to have only a few constraints to work with. The small number of constraints made it impossible to extract useful information about the mass distribution of the lens without using at least a simple parametrization. However, the quality and quantity of lensing data are growing fast, motivating the use of algorithms

capable of exploiting the available information with fewer assumptions. In recent years it has been possible to identify hundred of strong lensing arcs around the central region of galaxy clusters (Broadhurst et al., 2005). This drives us to the non-parametric (or free-form) approach, where the surface mass density of the cluster is treated as a combination of individual density functionals. Non-parametric approaches have been previously explored in several papers (Saha and Williams, 1997; Abdelsalam et al., 1998a,b; Trotter et al., 2000; Warren and Dye, 2003; Bradač et al., 2005; Liesenborgs et al., 2007). Since the resulting systems of equations are not always determined, non-parametric models are sadly not exempt of requiring some assumptions. In the following chapter we present the analysis of a galaxy cluster using a hybrid method, in the sense that it relies on the decomposition of the mass distribution in several basis functions, but also makes the valid assumption that a mass component traces the light of the galaxies within the cluster. For this reason, the details of the mass decomposition will be discussed further in Sec. 4.3.

Chapter 4

Strong lensing analysis of galaxy cluster MACS J1206.2-0847

Galaxy clusters are the largest gravitationally bound structures in the Universe, and a reflect of the late evolution of cosmic structures (see Schneider, 2006, for a review). Even though they were identified at first as regions with a significant overdensity of galaxies, it is now known that they consist of mainly three different mass components. In first place, we have the visible galaxies and stars. Furthermore, there is the *intracluster medium* (ICM), a thermal plasma mostly made up of ionized hydrogen and helium. The ICM accounts for most of the baryonic mass in galaxy clusters, and is detected through its strong X-ray emission. The lacking mass is distributed forming a dark matter halo, which constitutes by far the largest fraction of the total mass within the cluster.

Constraining the mass profile in the most central area of galaxy clusters is of particular interest, since this high-density region contains the largest densities of dark matter. These measurements provide good tests of dark matter models. If dark matter has a small cross section for interaction, or if dark matter is ultralight (e.g. in axion-like-particle, or ALP, models), it is at these regions where deviations from pure collisionless cold dark matter models are expected (Rocha et al., 2012; Kaplinghat et al., 2013, 2014). Also, if dark matter annihilates or interacts with itself, the density in the central region of galaxy clusters is expected to be sensitive to these interactions. In particular, if the cross section for self-interaction is significantly larger than 1 $\mathrm{cm}^2\mathrm{g}^{-1}$, the profile should flatten in the most central region, and offsets between the positions of the brightest cluster galaxy (BCG) and the peak of the dark matter distribution are expected if the cluster is in a merger phase. The central mass profiles of galaxy clusters can be inferred in several ways, each one probing a different range of radii. These methods include stellar kinematics, strong

lensing, weak lensing, Sunyaev-Zel'dovich, and X-ray emission, which cover the distance range from 10 kpc to 1 Mpc scale (Newman et al., 2009; Umetsu et al., 2011; Hogan et al., 2017; Andrade et al., 2019). Strong lensing methods are based on the magnification, shape, and positions of the multiple lensed images of background sources, which are used to set constraints on the mass distribution of the deflector. These images typically appear near the Einstein radius of the lens, or tangential critical curve, which is typically between tens and a few hundred kiloparsecs in galaxy clusters. Arcs can also appear close to the radial critical curve, which can be much closer (a few kiloparsecs) to the center of the cluster.

The interpretation of strong lensing observations in galaxy clusters can be difficult due to the distortions introduced in the background galaxies and also the lack of information (such as redshifts, Blandford and Narayan, 1992; Schneider et al., 1992; Wambsganss, 1998; Narayan and Bartelmann, 1996; Kneib, 2002). Fortunately, in recent years, high quality data has enabled us to spectroscopically confirm many families of multiple lensed images, reducing the level of uncertainty and allowing detailed lens reconstructions in many galaxy clusters. Lens reconstruction methods are broadly classified into parametric and free-form methods. Parametric models are the natural choice for modeling strong lensing clusters when the number of lensing constraints is relatively small, such as in the case of QSO lensing, where only three or four multiple images are typically available. These models require that initial assumptions, or priors, about the cluster mass distribution be made. Some of these assumptions are, for instance, that the dark matter halos follow the luminous matter in the cluster or that galaxy profiles have certain symmetries. However, if the number of lensing constraints is sufficiently high, e.g. a few tens to a hundred, it is possible to accurately reconstruct the mass profile of a galaxy cluster, including substructure, in a free-form way. The advantage of a free-form method is that no initial assumptions about the distribution of mass in the galaxy cluster are needed. Lens models derived with free-form methods are useful in these cases where the geometry of the lens is complex, such as in the galaxy cluster MACS0717. In this case, a free-form lens model was able to correctly predict the position of new families of lensed galaxies, which were later confirmed spectroscopically (Diego et al., 2015). Even in cases where the cluster is more regular, free-form methods have proven to be accurate at predicting different observations that were later confirmed. For instance, they predicted the time and position of the reappearance of SN Refsdal correctly (Diego et al., 2016), or the redshift of the family of lensed images in system 7 in the cluster A370 (Diego et al., 2018a).

In this chapter we explored, with strong lensing data alone, the galaxy cluster MACS J1206.2-0847 (hereafter J1206) using the free-form code WSLAP+ (Diego et al., 2005a,b, 2007; Sendra et al., 2014) aiming at estimating its inner total mass distribution (baryonic and dark matter) and constraining the dark matter component.

4.1 The galaxy cluster MACS J1206.2–0847

The cluster J1206 is at $z = 0.439$, and was originally discovered in a short 2-min r-band image taken on June 15th 1999 by the University of Hawaii's 2.2-meter telescope (UH2.2m) on Mauna Kea. This observation was triggered by the presence of a X-ray source previously detected by the ROSAT All-Sky Survey (RASS) (RXC J1206.2–0848, Böhringer et al. 2001, 2004). After its discovery, it was included in the MAssive Cluster Survey (MACS), an ongoing project aimed at the compilation of a statistically complete sample of very X-ray luminous (and thus, by inference, massive), distant clusters of galaxies from RASS data (Ebeling et al., 2001, 2007, 2010). The cluster appeared fairly relaxed in its X-ray emission, but had significant amounts of intracluster light which was not centrally concentrated, suggesting that galaxy-scale interactions were still ongoing despite the overall relaxed state (Eichner et al., 2013). The BCG can be easily identified and is located in the center of the cluster.

Following its initial identification as a potentially massive galaxy cluster, several follow-up observations were performed to confirm its cluster nature and characterise its physical properties. This cluster is actually one of the 25 clusters in the Cluster Lensing and Supernova survey with Hubble (CLASH) program (Postman et al., 2012), a major project aimed to build a large strong lensing data set on massive relaxed lensing clusters (Zitrin et al., 2012; Umetsu et al., 2012; Biviano et al., 2013). One of the most peculiar lensed images in this cluster is a 15-arcsec long, bright and tangential gravitational arc, extending from a few arcsec to ≈ 20 arcsec west of the BCG. This arc was spectroscopically observed first by Sand et al. (2004) and later confirmed, with a counter-image, by Ebeling et al. (2009) at $z = 1.036$. Ebeling et al. (2009) make an analysis of the total cluster mass distribution as well as of the mass in the cluster core using X-ray, dynamical and strong lensing data. This giant arc and its companion image are treated as a seven-image multiple system in order to perform a strong lensing model. Their strong lensing analysis, based on imaging from the *Hubble* Space Telescope (HST), yields a mass estimate twice larger than the X-ray mass estimate based on *Chandra* data. Despite its relaxed appearance at optical and X-ray wavelengths, this discrepancy is believed to be indicative of ongoing merging activity, which is also supported by a small, but significant, offset of the X-ray emission peak from the BCG, and the very high dispersion velocity of the cluster. Zitrin et al. (2012) exploit UV, optical and IR data from the 16 HST bands in the CLASH project along with spectra from VLT/VIMOS in order to spectroscopically identify 47 new multiple lensed images of 12 distant sources. A total of 50 multiple lensed images of 13 sources, spanning redshifts from 1.0 to 5.5, were known by then. They use 32 secure multiple lensed images belonging to nine sources in a parametric strong lensing reconstruction to derive a detailed

mass distribution, and to constrain the inner mass profile of this cluster.

A later comprehensive analysis of this cluster combining weak and strong lensing data from wide-field Subaru imaging and HST observations was performed by Umetsu et al. (2012). They primarly use the parametric strong lens modeling of Zitrin et al. (2012) but also perform some semi-independent free-form strong lensing analyses. They find good agreement in the overlap region between weak and strong lensing constraints. Agreement is found also between the lensing results and X-ray mass measurements, again in the overlap region of the two data sets. Their morphological analysis of both the distribution of galaxies and the reconstructed mass distribution also reveals the presence of an elongated large-scale structure surrounding J1206. This structure runs approximately NW-SE, and it is aligned with the cluster and BCG orientations. The overall shape of J1206 indicates that light follows the mass up to the large scales of the cosmic web probed by observations.

A conclusive assessment on the degree of relaxation of the cluster was finally reached in Biviano et al. (2013) through the analysis of the velocities of several hundreds of cluster members. Building on a data set of ~ 600 spectroscopic redshifts of cluster member galaxies out to the cluster outskirts (~ 5 Mpc), they perform a mass profile reconstruction with galaxy dynamics well beyond the cluster virial radius. The mass profile obtained is found to be in excellent agreement with those derived from all the other independent probes (that is, SL, WL and X-rays) and consistent with a NFW profile. This agreement lastly indicates that possible systematic biases in dynamical analyses have been properly accounted for, and that the cluster is in fact in a relaxed dynamical state. A pretty good agreement between *Chandra* X-ray and lensing mass profiles for this cluster is also reported in Donahue et al. (2014). In this case, they use a free-form analysis which combines weak and strong lensing constraints from the HST and from ground-based wide-field data (Merten et al., 2009, 2015).

Finally, the most up-to-date strong lensing analysis of this cluster is performed by Caminha et al. (2017). They were able to identify 23 new spectroscopically confirmed multiple lensed sources, using deep VLT/MUSE (Multi Unit Spectroscopic Explorer) observations in combination with CLASH-VLT imaging (Biviano et al., 2013). Using the positional measurements of 82 spectroscopic multiple images belonging to 27 families and a flexible parametric approach, they present an improved measurement of the mass distribution in the inner region of J1206. An extension of this strong lensing analysis, including stellar kinematics constraints to model the cluster galaxy population, and using the same catalog of multiple images and cluster galaxies as in Caminha et al. (2017), is described in Bergamini et al. (2019).

4.2 Lensing data

We used public imaging data obtained from the ACS and WFC3 Hubble instruments. We retrieved the data from the Mikulski Archive for Space Telescope[1] (MAST). For the optical data we used filters F275W, F336W, F390W, F475W, F606W, F775W, F814W and F850LP (ID 12069, PI: M. Postman) and for the IR data we used filters F105W, F110W, F125W, F140W and F160W (ID 12069, PI: M. Postman). We produced color images by combining these optical and IR bands.

Regarding the member galaxies in J1206, we restricted ourselves to only those galaxies that have been spectroscopically confirmed. As part of the CLASH-VLT Large Programme, VLT/VIMOS was used to measure redshifts for a sample of 2,736 sources over an area of 400 arcmin^2, spectroscopically confirming 577 galaxy members (Biviano et al., 2013; Rosati et al., 2014)[2]. They are defined as the galaxies with a velocity dispersion lower than $\Delta_v = 2919$ km s^{-1} in the cluster rest frame ($z = 0.439$, Girardi et al. 2015), which corresponds to a redshift range of $0.425 \leq z \leq 0.453$. Considering among them only those member galaxies that fall within the HST field of view and also are closer to the BCG, we selected the brightest ones and those which are near to strong lensing arcs. The lensing reconstruction method used is only sensitive to galaxies which are really close to an arc (a few arcsec) or that are very massive, and thus its effective radius of influence is larger than a few arcsec. Therefore, only those galaxies which meet these criteria are key for the lensing reconstruction method. The 54 cluster members selected (including the BCG) are listed in Table 4.1, and the closest ones to the center of the cluster are shown in Fig. 4.1.

For the strong lensing data, we followed the multiple-image system identifications from Caminha et al. (2017), which include 82 spectroscopic multiple images belonging to 27 background sources in the redshift range from 1.01 to 6.06. They combine spectroscopic measurements from CLASH-VLT (Biviano et al., 2013) and MUSE to find the largest possible number of families of multiple images based on their redshift values. Only multiple images of background galaxies with secure spectroscopic confirmation (identification in the MUSE and HST data) are considered in order to avoid any image misidentification. Multiple images located very close in projection to massive spectroscopically confirmed background galaxies are also discarded because these member galaxies can introduce significant deflections (difficult to model) in addition to those associated with the cluster. From Caminha et al. (2017), we also adopted the numbering scheme (except for some systems we decided to redefine in this work or the systems that include multiple images that

[1] https://archive.stsci.edu/prepds/clash/

[2] This spectroscopic catalog was publicly released in March 2014 and is available at https://archive.stsci.edu/missions/hlsp/clash/macs1206/catalogs/redshift.

were not used for the reasons explained above) as well as the spectroscopic redshifts. In addition to the centroid positions of the multiple lensed systems, we also used the position of individual knots present in well-resolved elongated arcs, that are easily identified thanks to the depth of HST data. The addition of extra knots in well-resolved systems has proven to greatly improve the accuracy and stability of the derived lensing solutions of the WSLAP+ method (Diego et al., 2016) due to the large extension of the giant arcs. We redefined the 15-arcsec long tangential arc at 20 arcsec west of the BCG (ID = 2b) by replacing the original two positions, 2b and 2c, (placed in both extremes of this arc) by ten constraints, more or less equally spaced, along the arc. Adding these additional constraints allows us to include information on the magnification of giant arcs as the optimisation algorithm tries to focus all knots into a very compact region in the source plane. It is important to emphasize that, given the small physical size of the sources, we approximated them at first by point sources, so each knot maps to the same location in the source plane. The shape of sources with multiple lensed knots, that are extended in the source plane, is reconstructed through the iterative process in the WSLAP+ method. A first model is run in which the spatial information of sources is not included and fewer iterations are done. Then a second model is run, where the deflection angle map obtained from the first model is used to predict the configuration of the knots in the source plane. The relative position of the knots in the source plane with respect to the centroid of the source is then fixed, so only the position of the centroid is a free parameter in the reconstruction. This was done several times increasing the number of iterations at each step until the process ends converging. The shape of extended sources in the source plane was reconstructed by this procedure, avoiding an undesired bias to models which might favour high magnifications due to the initial assumption that the sources are point-like. This technique is used in Diego et al. (2016) for source 1, which has as many as 20 well resolved knots. The configuration of the knots for source 1 is similar to those found by other methods (including parametric ones), suggesting that this iterative process, together with its convergence, is robust against different modeling methodologies. We also redefined the straight arc pointing NW (ID = 4b) that arise from the BCG by splitting it into five different knots spread along the arc. On the other hand, counter-images 7c and 7d, which are 3 arcsec apart, were considered as knots belonging to a unique curved arc leading NE (ID = 7c). Three other knots, adding to a total of five, were included in this curved counter-image. Accordingly, counter-image 7e was renamed as 7d. In summary, we use 97 lensing constraints in this work. All these constraints are listed in Table 4.2, and the nearest ones to the BCG are shown in Fig. 4.2.

In this work we adopt a flat cosmological model with $\Omega_M = 0.3$, $\Lambda = 0.7$, and $h = 70$ $\mathrm{Km\,s^{-1}Mpc}$. For this model, 1 arcsec = 5.68 kpc at the distance of the cluster ($z = 0.439$). The reference point of our system of coordinates is the BCG: RA= 12:06:12.1436, Dec = -8:48:03.055 (J2000). In all images (except when noted otherwise), we adopt the standard convention where north is up and east is left.

N°	RA	Dec	N°	RA	Dec
1	12:06:08.0928	-8:48:23.821	28	12:06:12.3353	-8:47:59.064
2	12:06:08.5999	-8:48:12.416	29	12:06:12.5191	-8:47:36.737
3	12:06:08.6263	-8:47:54.737	30	12:06:12.5676	-8:47:39.793
4	12:06:08.6426	-8:48:26.946	31	12:06:12.6288	-8:48:52.153
5	12:06:08.7348	-8:47:33.508	32	12:06:12.7015	-8:48:04.831
6	12:06:09.1373	-8:47:23.132	33	12:06:12.7282	-8:46:48.205
7	12:06:09.3401	-8:47:21.786	34	12:06:12.7418	-8:47:41.503
8	12:06:09.8885	-8:47:32.024	35	12:06:12.7493	-8:48:52.625
9	12:06:09.9206	-8:48:27.414	36	12:06:12.7822	-8:48:05.627
10	12:06:10.3814	-8:47:55.388	37	12:06:12.9187	-8:48:00.443
11	12:06:10.6385	-8:48:56.372	38	12:06:13.2612	-8:48:21.964
12	12:06:10.7038	-8:47:57.444	39	12:06:13.2898	-8:47:36.773
13	12:06:10.7573	-8:48:09.302	40	12:06:13.6320	-8:48:38.160
14	12:06:10.7623	-8:48:02.588	41	12:06:13.7657	-8:47:46.172
15	12:06:10.7856	-8:47:53.322	42	12:06:13.7846	-8:48:08.730
16	12:06:10.8290	-8:48:06.570	43	12:06:14.2651	-8:47:26.182
17	12:06:10.9505	-8:47:58.855	44	12:06:14.5106	-8:48:46.404
18	12:06:10.9543	-8:47:12.203	45	12:06:15.0310	-8:48:37.505
19	12:06:11.3618	-8:48:22.100	46	12:06:15.0482	-8:47:55.752
20	12:06:11.4444	-8:48:56.200	47	12:06:15.3494	-8:47:43.267
21	12:06:11.5236	-8:47:54.982	48	12:06:15.3562	-8:47:43.440
22	12:06:11.5370	-8:48:19.361	49	12:06:15.6612	-8:48:21.823
23	12:06:11.7094	-8:47:14.748	50	12:06:15.7313	-8:48:12.200
24	12:06:11.7270	-8:47:48.536	51	12:06:15.9929	-8:48:17.233
25	12:06:12.0842	-8:47:31.330	52	12:06:16.0973	-8:48:06.260
26	12:06:12.1459	-8:48:03.316	53	12:06:16.0990	-8:48:11.948
27	12:06:12.2818	-8:47:55.864	54	12:06:16.2806	-8:48:02.956

Table 4.1: Cluster member galaxies in J1206 used in our strong lensing analysis.

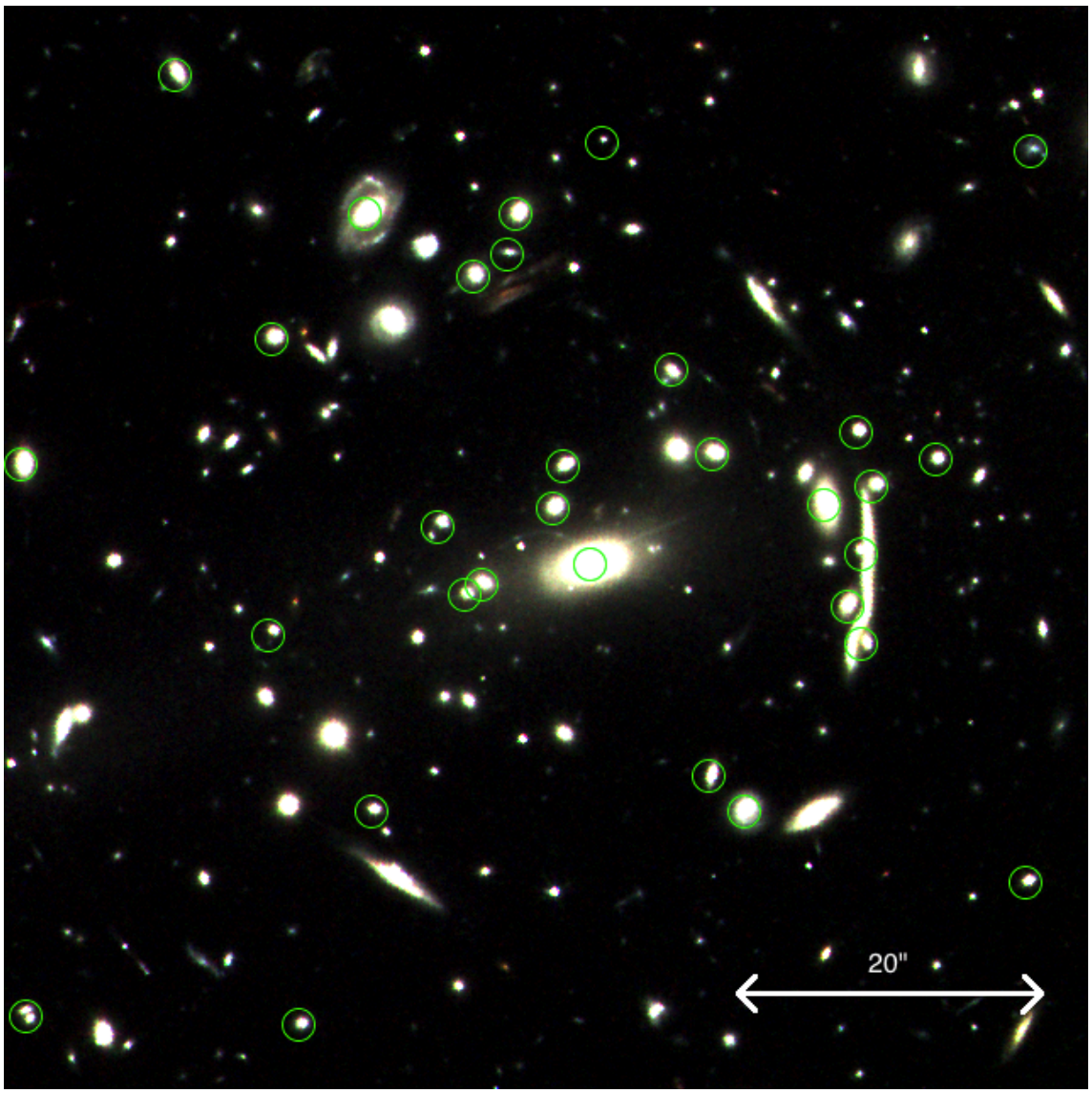


Figure 4.1: Image of cluster J1206 showing 31 of the 54 cluster members used in our analysis. The member galaxies that fall within a field of view of 1.18 arcmin centered in the BCG, are marked with green circles.

ID	RA	Dec	z_s
1a	12:06:12.2198	-8:47:50.7192	1.0121
1b	12:06:11.9050	-8:47:57.4584	1.0121
1c	12:06:11.7288	-8:48:23.9580	1.0121
2a	12:06:11.2296	-8:47:44.4480	1.0369
$2b_1$	12:06:10.7420	-8:47:58.1560	1.0369
$2b_2$	12:06:10.7396	-8:47:58.9030	1.0369
$2b_3$	12:06:10.7358	-8:47:59.4920	1.0369
$2b_4$	12:06:10.7304	-8:48:00.1940	1.0369
$2b_5$	12:06:10.7268	-8:48:01.9390	1.0369
$2b_6$	12:06:10.7358	-8:48:03.4150	1.0369
$2b_7$	12:06:10.7322	-8:48:05.2050	1.0369
$2b_8$	12:06:10.7541	-8:48:07.3880	1.0369
$2b_9$	12:06:10.8112	-8:48:09.5570	1.0369
$2b_{10}$	12:06:10.8245	-8:48:10.9872	1.0369
3a	12:06:12.1368	-8:47:44.0448	1.0433
3b	12:06:11.4266	-8:47:59.3196	1.0433
3c	12:06:11.6657	-8:48:19.0116	1.0433
4a	12:06:12.7169	-8:47:40.9164	1.4248
$4b_1$	12:06:11.7192	-8:48:00.2052	1.4248
$4b_2$	12:06:11.8088	-8:48:00.6250	1.4248
$4b_3$	12:06:11.8837	-8:48:01.0500	1.4248
$4b_4$	12:06:11.9246	-8:48:01.2810	1.4248
$4b_5$	12:06:12.0083	-8:48:01.9110	1.4248
4c	12:06:11.9405	-8:48:28.6740	1.4248
5a	12:06:12.8537	-8:47:42.6804	1.4254
5b	12:06:13.0169	-8:48:05.5872	1.4254
5c	12:06:12.0012	-8:48:29.1528	1.4254
6a	12:06:11.9950	-8:47:46.9032	1.4255
6b	12:06:11.5534	-8:47:49.4088	1.4255
6c	12:06:11.5320	-8:48:33.4188	1.4255
7a	12:06:12.1351	-8:47:44.5344	1.4257
7b	12:06:11.3263	-8:47:52.7928	1.4257
$7c_1$	12:06:12.2328	-8:48:01.8410	1.4257
$7c_2$	12:06:12.3214	-8:48:01.1808	1.4257
$7c_3$	12:06:12.4327	-8:48:00.9640	1.4257
$7c_4$	12:06:12.5311	-8:48:01.4328	1.4257
$7c_5$	12:06:12.6092	-8:48:02.0260	1.4257
7d	12:06:11.5937	-8:48:31.8816	1.4257
8a	12:06:12.8777	-8:47:44.7216	1.4864
8b	12:06:13.0858	-8:48:03.9744	1.4864
8c	12:06:11.9897	-8:48:31.9932	1.4864
9a	12:06:11.2178	-8:47:35.3184	1.9600
9b	12:06:10.3855	-8:47:52.1232	1.9600
9c	12:06:10.6507	-8:48:26.9496	1.9600
10a	12:06:12.5880	-8:47:42.0036	2.5393
10b	12:06:11.1850	-8:47:50.8740	2.5393
10c	12:06:12.1169	-8:47:59.8452	-
10d	12:06:13.1746	-8:48:00.5760	-
10e	12:06:11.7185	-8:48:42.5268	-

ID	RA	Dec	z_s
11a	12:06:15.0370	-8:47:48.0192	3.0358
11b	12:06:14.9988	-8:48:17.6796	3.0358
11c	12:06:14.5375	-8:48:32.3568	3.0358
12a	12:06:11.6717	-8:47:37.3812	3.3890
12b	12:06:11.0690	-8:47:43.3932	3.3890
12c	12:06:12.7843	-8:48:00.7092	3.3890
13a	12:06:11.8598	-8:48:06.3648	3.3961
13b	12:06:11.4862	-8:48:07.6140	3.3961
13c	12:06:12.9048	-8:48:17.1324	3.3961
13d	12:06:12.3334	-8:48:24.7104	3.3961
14a	12:06:15.9739	-8:48:16.1280	3.7531
14b	12:06:15.9540	-8:48:17.0388	3.7531
14c	12:06:15.9540	-8:48:18.5292	3.7531
14d	12:06:15.9060	-8:48:22.7808	3.7531
14e	12:06:15.7418	-8:48:27.6840	3.7531
15a	12:06:13.4309	-8:47:29.8860	3.7611
15b	12:06:13.8240	-8:48:11.0016	3.7611
15c	12:06:12.4195	-8:48:39.4704	-
16a	12:06:13.1002	-8:47:28.3272	3.7617
16b	12:06:11.1516	-8:47:58.8156	3.7617
16c	12:06:13.5648	-8:48:08.8956	3.7617
17a	12:06:13.4726	-8:47:44.2320	3.8224
17b	12:06:13.6699	-8:47:57.9192	3.8224
18a	12:06:13.2902	-8:47:48.1704	4.0400
18b	12:06:13.4225	-8:47:54.9420	4.0400
19a	12:06:14.9002	-8:47:41.5500	4.0520
19b	12:06:14.8495	-8:48:18.8604	4.0520
19c	12:06:14.3491	-8:48:34.0668	4.0520
20a	12:06:11.3933	-8:48:01.7136	4.0553
20b	12:06:13.6414	-8:48:13.7268	4.0553
21a	12:06:10.4234	-8:47:51.6264	4.0718
21b	12:06:12.3509	-8:48:02.9304	4.0718
21c	12:06:12.6840	-8:48:04.0140	4.0718
21d	12:06:12.8232	-8:48:04.3344	4.0718
22a	12:06:10.6387	-8:47:29.1048	4.2913
22b	12:06:9.8122	-8:48:21.9384	4.2913
23a	12:06:15.1805	-8:47:48.8148	4.7293
23b	12:06:15.2489	-8:48:13.2120	4.7293
24a	12:06:12.3307	-8:47:28.6800	5.6984
24b	12:06:10.5804	-8:47:49.5384	5.6984
24c	12:06:12.2186	-8:48:02.1852	5.6984
24d	12:06:13.2504	-8:48:03.8376	5.6984
25a	12:06:14.3314	-8:47:47.6232	5.7927
25b	12:06:14.4245	-8:48:00.6372	5.7927
26a	12:06:12.1706	-8:48:11.2032	6.0106
26b	12:06:12.2906	-8:48:13.2048	6.0106
27a	12:06:12.7802	-8:47:55.7700	6.0601
27b	12:06:12.4058	-8:47:57.0012	6.0601

Table 4.2: Spectroscopic multiple images in J1206 used in our strong lensing analysis.

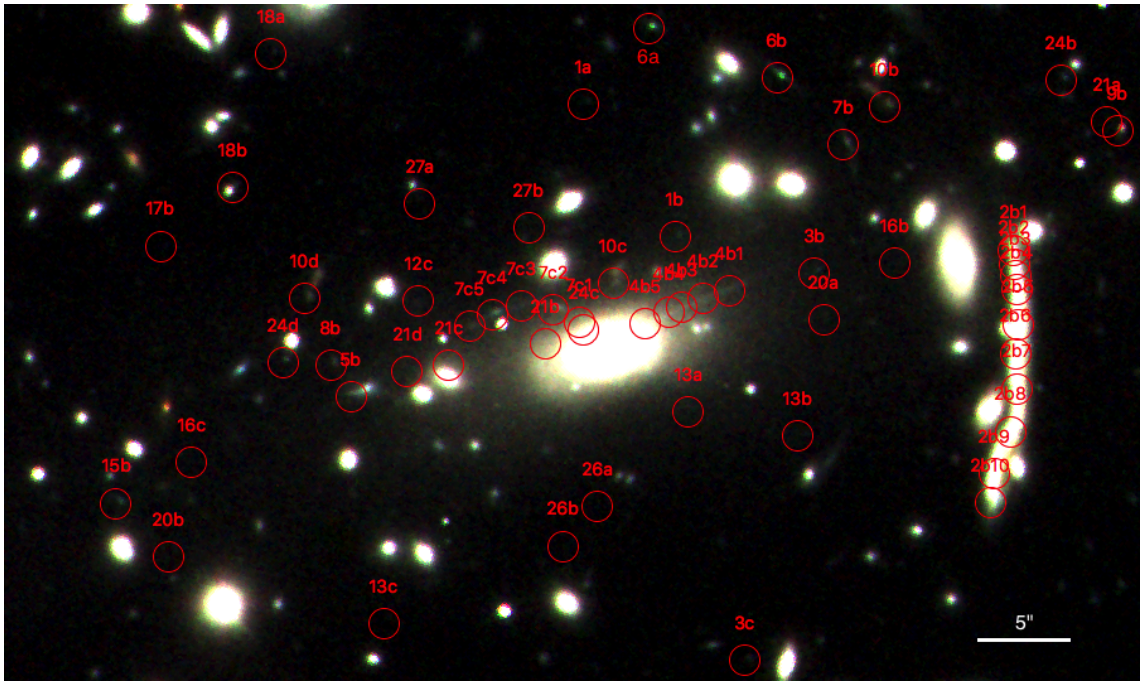


Figure 4.2: Image of cluster J1206, where all the spectroscopically confirmed multiple images used in our strong lensing models, and that fall within a field of view of 58 arcsec, are indicated with red circles.

4.3 Lensing reconstruction method

The lens model for J1206 was obtained using the WSLAP+ code. This algorithm does not rely on assumptions on the dark matter distribution and has proven to perform equally well both in morphologically complex clusters (Diego et al., 2015) as well as with more regular clusters (Diego et al., 2016, 2018a). This free-form method allows for increased spatial resolution, and flexibility, in certain areas of the lens plane, where more detail is needed. We took advantage of this feature to study the central region of this cluster in greater detail and benefit from the constraints provided by two radial arcs with unusually extended morphology and close to the center of the cluster. When the performance of WSLAP+ was compared with other codes using simulated observations (Meneghetti et al., 2017), the lens models derived with WSLAP+ reproduced well the lensing potential in those regions where multiple images are present. On the contrary, and not surprisingly, the potential was poorly constrained in those regions without lensing constraints. A description of the main aspects of this method is given in this section. Further details of the method can be found in Diego et al. (2005a,b, 2007) and Sendra et al. (2014).

WSLAP+ (Weak and Strong Lensing Analysis Package) falls in the category of hybrid methods as it relies on a grid configuration (free-form) plus a compact component that traces the light from the cluster galaxies. The code is able to estimate the cluster mass distribution without using any prior information on the underlying mass provided the number of strong lensing images with known redshifts is sufficiently large. No priors on the shape of the individual halos are needed. For the usually small compact component, mostly the baryonic contribution, the code assumes that the mass is proportional to the observed flux. All galaxies can be forced to have the same mass-to-light ratio or some galaxies (usually the BCG or other bright galaxies) are allowed to have their own mass-to-light ratio. This method relies on the decomposition of the lens plane into individual cells. We show here how the problem in this approximation can be expressed as a system of linear equations, for which a solution can be found.

The fundamental problem in lens modeling is to find out what are the positions $\boldsymbol{\beta} = (\beta_x, \beta_y)$ of the background sources and the mass distribution $m(\theta)$ of the lens given a series of lensed images whose positions $\boldsymbol{\theta} = (\theta_x, \theta_y)$ are known. This is achieved by solving the standard lens equation enunciated in Eq. (3.8):

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}, m(\boldsymbol{\theta})) \tag{4.1}$$

where $\boldsymbol{\alpha}$ is the deflection angle caused by the lens. The non-linear dependence of the deflection angle $\boldsymbol{\alpha}$ on the observed positions $\boldsymbol{\theta}$ could make us think we are facing a a non-linear problem. However, since the positions $\boldsymbol{\theta}$ are known through observation and the gravitational potential has a linear nature, a linear formulation of the lensing problem is possible. In the following lines we will describe how to reach this end.

The lensed images, or strong lensing observables $\boldsymbol{\theta}$, can be expressed in terms of derivatives of the lensing potential:

$$\psi(\boldsymbol{\theta}) = \frac{4G}{c^2}\frac{D_{LS}}{D_L D_S}\int m(\boldsymbol{\theta'})\ \ln(|\boldsymbol{\theta} - \boldsymbol{\theta'}|)\, d\theta' \tag{4.2}$$

where D_L, D_S and D_{LS} are the angular diameter distances to the galaxy cluster, to the background galaxies and from the cluster to the galaxies, respectively. To begin with, the gradient of this potential allows us to obtain the deflection angle:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{4G}{c^2}\frac{D_{LS}}{D_L D_S}\int m(\boldsymbol{\theta'})\frac{\boldsymbol{\theta} - \boldsymbol{\theta'}}{|\boldsymbol{\theta} - \boldsymbol{\theta'}|^2}\, d\theta' \tag{4.3}$$

If we assume the usual thin lens approximation, the mass distribution $m(\theta)$ of the lens

can be fully described by its surface mass density $\Sigma(\theta)$ within the defined lens plane. The expression of the deflection angle then takes the following form:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{4G}{c^2}\frac{D_L D_{LS}}{D_S}\int \Sigma(\boldsymbol{\theta'})\frac{\boldsymbol{\theta}-\boldsymbol{\theta'}}{|\boldsymbol{\theta}-\boldsymbol{\theta'}|^2}\, d^2\theta' \tag{4.4}$$

And according to Eq. (3.25), the magnification produced by the lens is easily derived from $\boldsymbol{\alpha}$ as

$$\mu^{-1}(\boldsymbol{\theta}) = 1 - \frac{\partial\alpha_x}{\partial x} - \frac{\partial\alpha_y}{\partial y} + \frac{\partial\alpha_x}{\partial x}\frac{\partial\alpha_y}{\partial y} - \frac{\partial\alpha_x}{\partial y}\frac{\partial\alpha_y}{\partial y} \tag{4.5}$$

The surface mass density on the lens plane can be described by the combination of two components:

- A soft (or diffuse) component, which we construct as the superposition of N_c basis functions on a grid split into as many cells as functions. The cells on this grid can have constant width (regular grid) or varying width (adaptive grid).
- A compact component, that accounts for the mass associated with the individual galaxies in the cluster.

Regarding the soft component, the projected mass, which must simultaneously satisfy the lens equation from Eq. (4.1) for all the images observed, can be expanded as a combination of individual functions:

$$\Sigma(\theta_x, \theta_y) = \sum_l c_l f_l(\theta_x, \theta_y) \tag{4.6}$$

where c_l are the corresponding coefficients for each one of the basis function $f_l(\theta_x, \theta_y)$ considered. The basis functions could, in principle, be any kind of two-dimensional function. For instance, orthogonal polynomials or wavelet functions. An study of how the lens models depend on the choice of basis function used to represent the mass distribution was carried out in Diego et al. 2007. From that discussion, it was concluded that the best results are obtained using compact basis functions defined on a lens plane split into small grids or cells. Extended basis functions tended to produce more images than were observed. Specifically, they decided to assume that the mass distribution in each cell follows a Gaussian distribution located at the center of the cell with a varying dispersion. Gaussian functions offer several advantages, including a good compromise between the desired compactness and smoothness of the basis function, and a fast analytical computation of the integrated mass for a given radius. Accordingly, here we will follow the same path and

use Gaussians for the f_l functions.

On the other hand, the compact component was modeled by adopting the light distribution of the cluster observed in the filter F160W, and directly assigning to each galaxy a mass proportional to its surface brightness. This mass is later re-adjusted in an optimisation process. The compact component is usually split into independent layers, each one containing one or several cluster members. The separation into layers allows us to constrain differently the mass associated with different kinds of galaxies. This is useful in the case where the light-to-mass ratio may be different, as happens for instance with the BCG.

If we substitute the mass decomposition from Eq. (4.6) in Eq. (4.4), the deflection angle for each background galaxy can be rewritten as

$$\boldsymbol{\alpha}(\theta_j) = \lambda_j \sum_l c_l \int f_l(\boldsymbol{\theta'}) \frac{\boldsymbol{\theta} - \boldsymbol{\theta'}}{|\boldsymbol{\theta} - \boldsymbol{\theta'}|^2} d\theta' = \lambda_j \sum_l c_l \tilde{f}_l(\theta_j) \tag{4.7}$$

where $\lambda_j = 4\mathrm{G}D_{LS}/c^2 D_S D_L$ is different for each background galaxy, since it contains all the distance factors whose values depend on the source. The factor $\tilde{f}_l(\theta_j)$ is the convolution of the basis function f_l with the kernel $(\boldsymbol{\theta} - \boldsymbol{\theta'})/|\boldsymbol{\theta} - \boldsymbol{\theta'}|^2$ at the position $\boldsymbol{\theta}$. If we now define the multiplication of these two variables as the matrix $\boldsymbol{\Upsilon}$ by:

$$\Upsilon_{jl} = \lambda_j \tilde{f}_l(\theta_j), \tag{4.8}$$

the lens equation shown in Eq. (4.1) can be expressed in the following form:

$$\boldsymbol{\Theta} = \boldsymbol{\Upsilon} \boldsymbol{c} - \boldsymbol{\beta} \tag{4.9}$$

Going one step further, if all the unknowns in the lens problem are grouped together, the strong lensing problem can finally be expressed as a system of linear equations, that can be depicted in an even more compact form:

$$\boldsymbol{\Theta} = \boldsymbol{\Gamma}\ \boldsymbol{X} \tag{4.10}$$

$$\begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Upsilon}_x & \mathbf{I}_x & \mathbf{0} \\ \boldsymbol{\Upsilon}_y & \mathbf{0} & \mathbf{I}_y \end{pmatrix} \begin{pmatrix} \boldsymbol{c} \\ \beta_x \\ \beta_y \end{pmatrix} \tag{4.11}$$

where the ij elements in matrix $\mathbf{I}_x$ are ones if the θ_i pixel (x-coordinate) comes from the

β_j source (y-coordinate), and zero otherwise. The matrix $\mathbf{I}_y$ is defined analogously, and the matrix $\mathbf{0}$ contains all zeros. In this system of equations, the measured strong lensing observables are contained in the array $\boldsymbol{\Theta}$ of dimension $N_\Theta = 2\ N_{SL}$, the unknown source positions β and surface mass density $\Sigma(\theta)$ are contained in the array $\boldsymbol{X}$ of dimension $N_X = N_c + N_g + 2N_s$, and the matrix $\boldsymbol{\Gamma}$, which accounts for all the physics and geometry of the grid, is known (for a given grid configuration and fiducial galaxy deflection field) and has dimension $N_\Theta \times N_X$. N_{SL} is the number of strong lensing observables (each one contributes with two constraints, θ_x, and θ_y). N_s is the number of background sources (each one contributes with two unknowns, β_x, and β_y), which in our case was fixed to 27 families spectroscopically confirmed with CLASH/VLT and MUSE data. And N_c is the number of grid points (or cells) that we use to split the field of view in the soft component of the surface mass density. A Gaussian function is placed in each grid point. The width of these Gaussians is chosen in such a way that two nearby grid points with a same amplitude result in a horizontal plateau between the two overlapping Gaussians.

In this work, we considered both regular and adaptive (multiresolution) grid configurations. Adaptive grids are useful when there is a clear peak in the mass distribution, for instance here where the cluster has a single well-defined BCG. The regular grid used in this work has $N_c = 32 \times 32 = 1024$ grid points. The multiresolution grid used in this work has $N_c = 480$ grid points. These two grids are shown in Fig. 4.3. N_g is the number of deflection fields (from cluster members) that we consider, that is, the number of mass layers in which we split the compact component. In this work, we set N_g equal to 1 or 2 in order to explore different configurations. We contemplated a scenario with $N_g = 2$ where all galaxies, except for the BCG, are in the same layer. On the other hand, when $N_g = 1$, all the galaxies in the cluster are re-scaled by the same parameter because they all are in the same layer. We contemplated four cases with $N_g = 1$: i) the first one in which all the individual galaxies were assumed to follow the same light-to-mass ratio, ii) a second case in which the light contribution of the BCG was decreased by a factor of 0.5, iii) a third case where the light distribution of the BCG was reduced to one third of its original weight, and iv) the last case in which the BCG was excluded and only the rest of member galaxies are taken into account. Cases ii-iv) are interesting to explore the possibility that the dark matter content within the BCG is very small and the baryons in the BCG can hide this deficit. We will return to the cases studied at the end of this section.

In summary, the strong lensing problem has been formulated in Eq. (4.10) in a manner where the observables Θ depend linearly on the unknowns X. Even though an exact solution for X exists as long as the inverse of Γ exists, Γ is singular in most cases and thus does not have a inverse. In consecuense, a direct inversion of the lensing problem is not always feasible. This seeming obstacle leads us to need a numerical approach to solve the

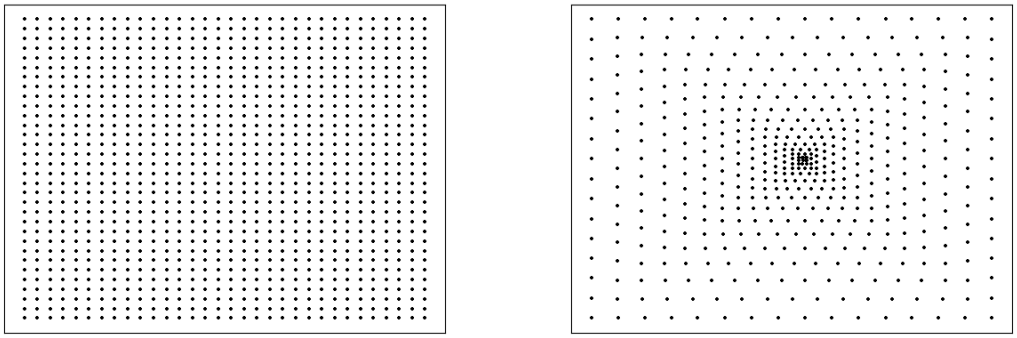


Figure 4.3: Depiction of the grids used in this work: a regular grid of 1024 cells (*left*) and a multiresolution grid of 480 cells (*right*). The peaks of the individual Gaussians are located at the possitions of the dots.

lens equation, which in the end turns out not to be as bad as you might think. For the cases where the inverse of Γ exists, we would obtain an exact solution of X, but taking into account the following assumptions made we would be more interested in slightly more imprecise solutions. We defined X assuming that the background galaxies are only placed at the coordiantes β_x and β_y, which is inaccurate as galaxies have some spatial extent. In addition, the approximation that the mass can be described as a superposition of basis functions is not fully accurate. Therefore, in order to allow for some looseness in the solutions derived, we incorporate the following residual ($|\boldsymbol{r}| > 0$) in the lens equation:

$$\boldsymbol{r} \equiv \boldsymbol{\Theta} - \boldsymbol{\Gamma}\,\boldsymbol{X} \tag{4.12}$$

that we assume is distributed following a Gaussian. The solutions we are looking for are those that maximize the following likelihood function:

$$\mathcal{L}(\boldsymbol{X}) = e^{-\frac{1}{2}\chi^2} \tag{4.13}$$

This is the same as saying they have to minimize a χ^2 we define as

$$\chi^2 = \frac{1}{2}\boldsymbol{X}^t\mathbf{A}\boldsymbol{X} + \boldsymbol{a}^t\boldsymbol{X} \tag{4.14}$$

where $a = 2\boldsymbol{\Gamma}^t\mathbf{C}^{-1}\boldsymbol{\Theta}$ and $\mathrm{A} = 2\boldsymbol{\Gamma}^t\mathbf{C}^{-1}\boldsymbol{\Gamma}$, being $\mathbf{C}$ the covariance matrix of the residual $\boldsymbol{r}$.

The solution of the linear system of equations shown in Eq. (4.10) is therefore found

after minimizing the above quadratic function of X. This minimization is done with the constraint that the solution, X, must be positive. Since the vector X contains the masses, the renormalization factors for the galaxy deflection field, and the background source positions, and all these quantities are always positive (the reference zero of the source positions is defined in the bottom left corner of the field of view), imposing $X > 0$ helps to constrain the space of meaningful solutions, and to regularise the solution as it avoids large negative and positive contiguous fluctuations. The approach followed here is known as multiplicative updates for non-negative quadratic programming, and was developed by Sha et al. (2003). The quadratic algorithm convergence is fast (80 minutes using the grid with $N_c = 1024$ in a 2.7 GHz Intel Core i7 processor) allowing for multiple solutions to be explored in a relatively short time. Different solutions can be obtained after modifying the starting point (or grid configuration) in the optimisation and/or the redshifts of the systems without spectroscopic redshift. A detailed description of the quadratic algorithm used in the WSLAP+ method can be found in Diego et al. (2007), and a recent discussion about its convergence and performance based on simulated data can be found in Sendra et al. (2014).

In order to account for uncertainties and variability in the lensing models, we explored a range of configurations where we changed the assumptions for the two main components of our method: the light-to-mass ratio of the member galaxies (compact component) and the grid definition (soft component). The number of member galaxies considered was always fixed to 54. We considered nine types of models, which can be grouped into five categories or cases, according to how the compact component was treated. In case 1, all galaxies but the BCG were in the same layer ($N_g = 2$). All galaxies were assumed to follow a fixed light-to-mass ratio while the BCG was allowed to have its own independent light-to-mass ratio. A multiresolution grid with $N_c = 480$ cells (1a) and a regular grid of $N_c = 1024$ cells (1b) were considered. From case 2 on, all the galaxies were in the same layer ($N_g = 1$). Such as in case 1, grids with 480 (2a, 3a, 4a) and 1024 cells (2b, 3b, 4b) were again used in cases 2, 3, and 4. In case 2, galaxies were forced to have the same light-to-mass ratio. In case 3, all galaxies but the BCG had the same light-to-mass ratio equal to one. The BCG was forced to have a mass-to-light ratio of 1/2 times the one for the other member galaxies. Case 4 is similar to case 3 but this time the BCG was forced to have a mass-to-light ratio of 1/3 times the one for the other member galaxies. Finally, in case 5 the BCG was excluded from the compact component and only the grid component contributed with mass in the central region. The other member galaxies had all the same light-to-mass ratio. Only results for the grid with 480 cells (5a) were used in this work[3].

[3]Regular grid with 1024 cells was also considered but we decided to exclude it due to its clear poor reconstruction of radial arc 7c.

4.4 Results

We derived a solution for each one of the nine models discussed at the end of Sec. 4.3. Solutions typically converge to a stable point after $\approx 10^5$ iterations, at which point the minimization was stopped. We show the critical curves for the two models in case 2, considering a source at redshift $z_s = 3$, in Fig. 4.4. Since most of the differences are due to changes in the grid configuration, the critical curves for the other cases look similar.

To test the quality of the solutions, we compared the observed and predicted counter-images in the image plane. First, we have focused our attention on system 2 since this was the first one spectroscopically observed due to the size and brightness of image 2b. We used image 2a to predict the entire system 2, by delensing 2a to the source plane and later relensing it back to the image plane. This was done for every lens model considered. The predictions of counter-image 2b for all nine models are shown in Fig. 4.5. The agreement between the observed and predicted images is in general good. All predicted images of this arc have the same length as the original one varying a bit in the thickness, straightness and the meandering in the lower region. The fact that all models with different configurations perform well at predicting this counter-image probes the robustness of the models. Cases 3b and 4b, and to a lesser degree 1b and 2b, are distinctively narrower in the middle section of the arc than the other cases (and the real data), indicating that the predicted radial magnification ($\mu_r = (1 - \kappa + \gamma)^{-1}$) in this region of the lens plane is smaller for these models. Interestingly, b-type models have a larger contribution from the member galaxies (compact component). The increase in mass in the b-type models is expected since these models correspond to the regular grid, which has a lower resolution around the central region. The impact of a lower resolution grid around the BCG is evident in case 1b, in which the mass of the BCG (M_2 in Table 4.3) is ≈ 3 times larger than in case 1a. Since b-type models have a larger κ around the member galaxies, the value of the radial magnification can be smaller near these galaxies than in a-type models.

Despite the robust prediction of counter-image 2b offered by all models considered, differences arise when looking at their projected density mass profiles. Differences between their profiles are logically greater the closer we approach the center of the cluster since the very central region is not constrained by any observation. Cases 1 and 2 have density profiles steeper in the inner region as they allow for a greater contribution from the BCG. On the other hand, the density mass profile from case 5, where the BCG was excluded, is shallower. Differences between different grid configurations with the same conditions for the compact component are less significant. We have estimated the values of the logarithmic slope of the projected total mass density profiles for all models within a few

kpc from the BCG. The total mass density profiles with the highest (case 1b) and lowest slopes (case 5a) are shown with black dashed-dotted and black dashed lines, respectively, in Fig. 4.6. The seven additional models fall in between these two models, and hence for clarity, are not shown in the plot.

The central part of J1206 is the most interesting region we aimed to study. The determination of the total mass distribution in the inner core of J1206 allows us to measure the dark matter density profile and test for possible models of dark matter, such as self-interacting dark matter, that predict shallow or even flat density profiles in the very central region extending up to tens of kpc (and difficult to explain with baryonic feedback). Since the very central region has a large contribution to the mass from the baryonic component itself, one needs to estimate, and later subtract, this baryonic contribution. We used 65-mas CLASH images to measure the photometry of the BCG. The CLASH program image each cluster with 16 HST filters from UV to near-IR wavelengths. However, due to insufficient overlap between individual exposures, certain filters are prone to cosmic ray contamination. Images from these filters were not included in our analysis. We only measured the photometry from F275W, F606W, F814W, F105W, F110W, F125W, F140W, and F160W images. The photometry of the BCG was measured by running SExtractor in dual-image mode. The F160W image was used for source detection, segmentation and the definition of the isophotal apertures. Fluxes and uncertainties in each band were then measured from these isophotal apertures.

We fitted synthetic stellar population spectra to the SED with FAST (Kriek et al., 2009). We made the following assumptions when constructing the synthetic spectra: (i) the Bruzual and Charlot (2003) stellar population synthesis model, (ii) the Chabrier (2003) initial mass function, (iii) an exponentially declining star formation history, (iv) the Kriek and Conroy (2013) dust law, (v) a fixed redshift of $z = 0.44$, and (vi) a solar metallicity of $Z = 0.02$. In addition to these assumptions, the synthetic populations are described by four free parameters, which include stellar mass, star formation timescale, age, and extinction. These free parameters are allowed to take on a wide range of physically plausible values. For each combination of their values, a synthetic spectrum was constructed. FAST then calculated the χ^2 of each synthetic spectrum to the observed SED, and found the best-fit spectrum, which is shown in Fig. 4.7. The stellar mass of the best-fit spectrum is $3.61^{+0.57}_{-1.98} \times 10^{11} M_{\odot}$ (3σ interval). In the end, after removing this stellar mass, we achieved an estimate of the dark matter distribution for each one of the models described at the end of Sec. 4.3. Dark matter density profiles for the models with the highest (case 1b) and lowest slopes (case 5a) are shown with red dashed-dotted and blue dashed lines, respectively, in Fig. 4.6. It is important to note how case 5a (which we remind it corresponds to the model where the contribution from the BCG is forced to be zero) results in a declining

dark matter density profile at distances below ≈ 3 kpc once the stellar mass is subtracted. This suggests that case 5a is underestimating the total mass at these radii. However, this is compensated by an excess of mass between ≈ 3 kpc and ≈ 10 kpc, keeping a total mass within 10 kpc approximately the same as for the other models. In fact, the integrated total mass within ≤ 10 kpc for case 5a ($6.43 \times 10^{12} M_{\odot}$) is slightly higher than the mass in the same region for the case 1b ($6.28 \times 10^{12} M_{\odot}$).

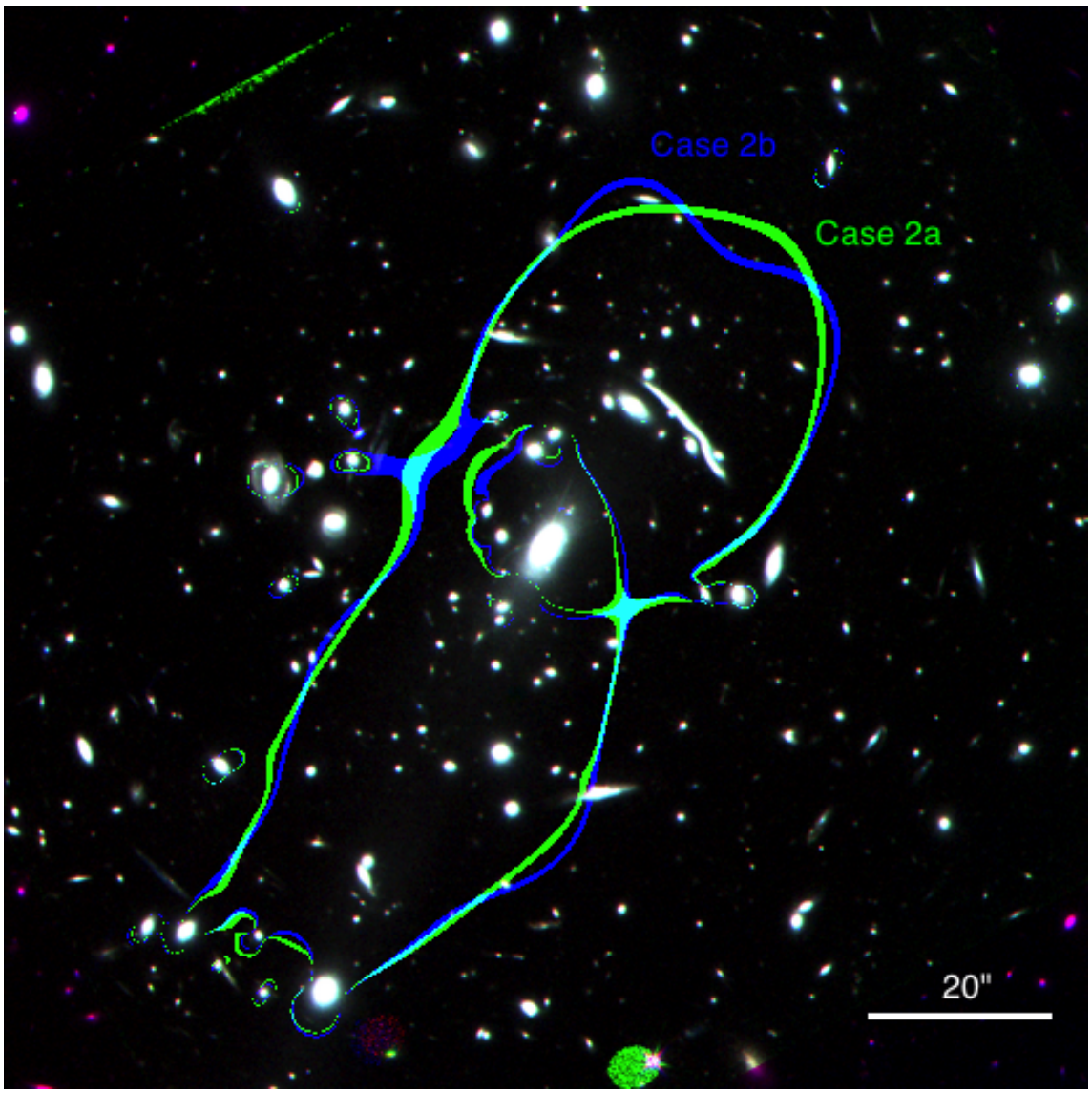


Figure 4.4: Critical curves for the two models in case 2 (case 2a in green and case 2b in blue), considering a source at $z_s = 3$. The field of view is 2 arcmin. The orientation is not north-south, but rotated 44° counterclockwise.

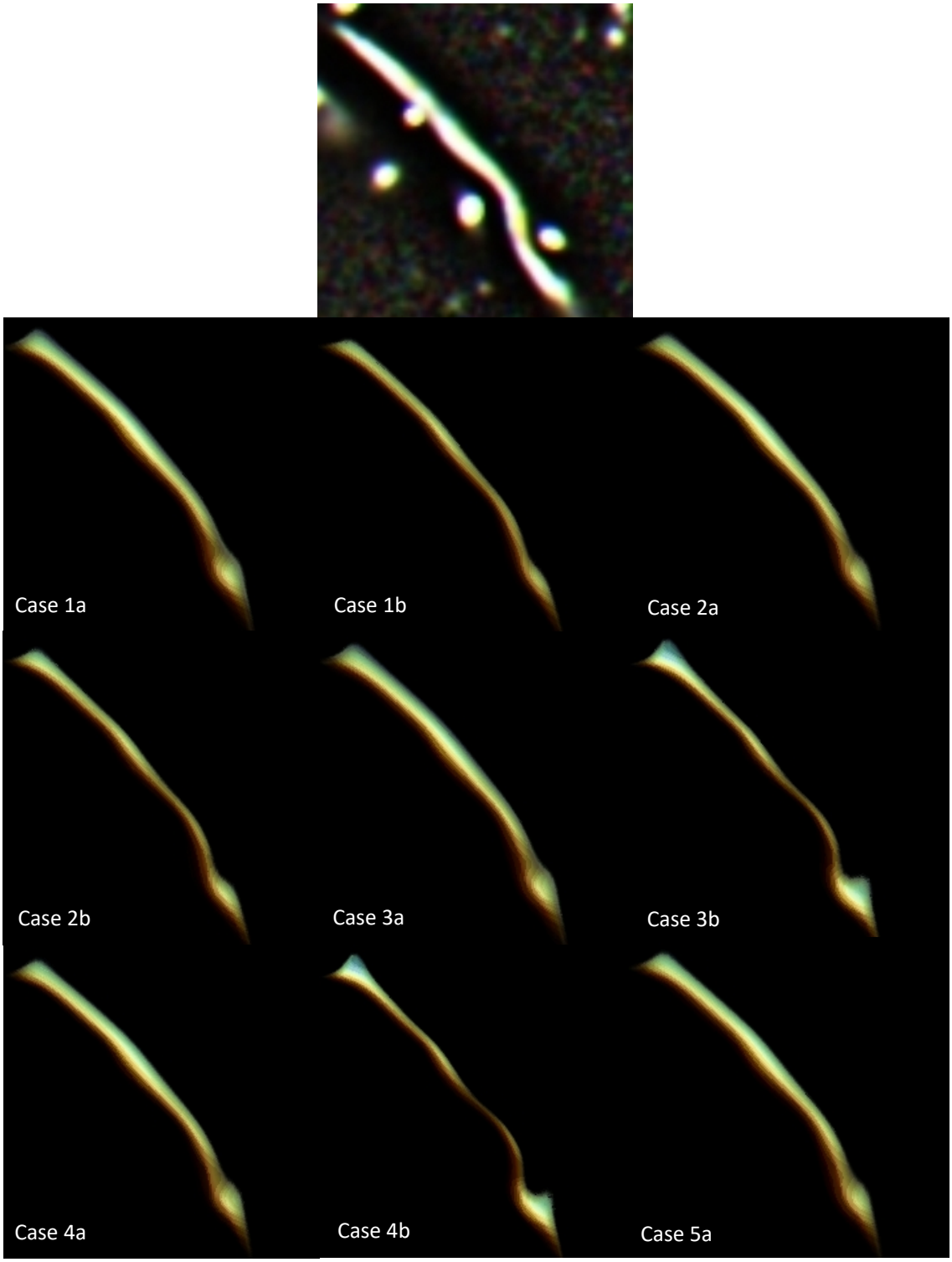


Figure 4.5: Original and predicted images of arc 2b, using the delensed counter-image 2a as a template of the source, for the nine different models discussed at the end of Sec. 4.3. The field of view and center of the images are the same in all panels. Image used was filtered to reduce light contribution from member galaxies of the cluster.

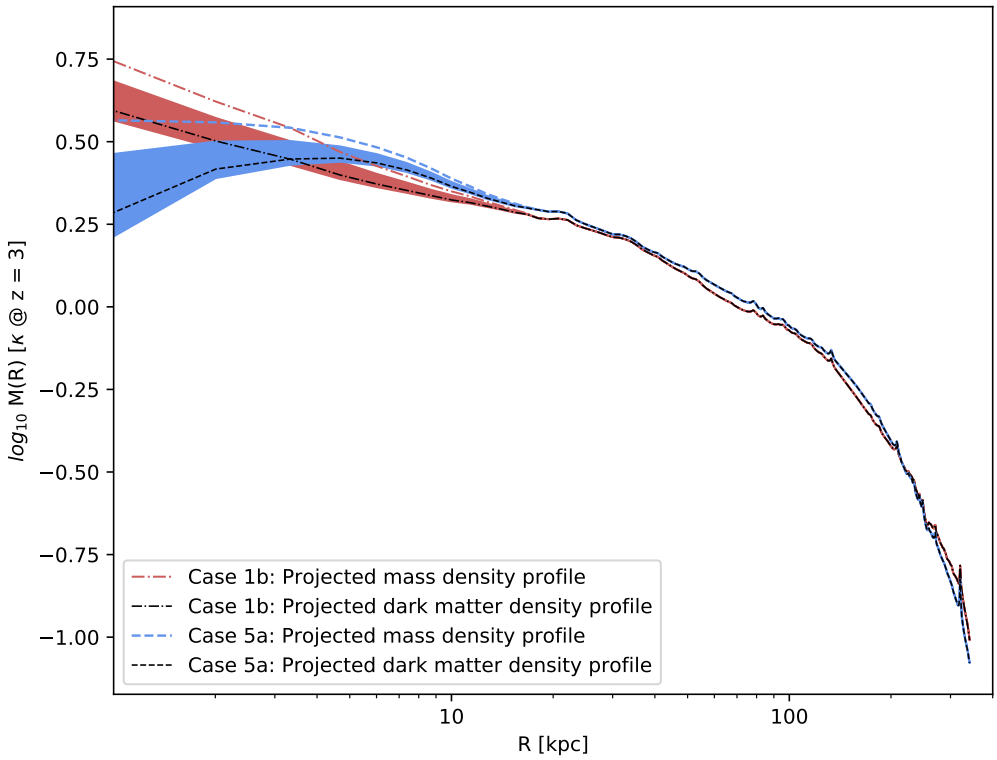


Figure 4.6: Mass profiles, in terms of the critical surface mass density (computed at $z_s = 3$), for the solutions with the highest (case 1b in red) and lowest slopes (case 5a in blue) within a few kpc from the center of the cluster. The red dash-dotted line and the blue dashed line are the total mass density profiles for each case. The black dash-dotted line surrounded by a red shaded uncertainty region (3σ) and the black dashed line surrounded by a blue shaded uncertainty region (3σ) are the dark matter density profiles in each case. Other models have not been included because their behavior falls within the borders established by the two models plotted.

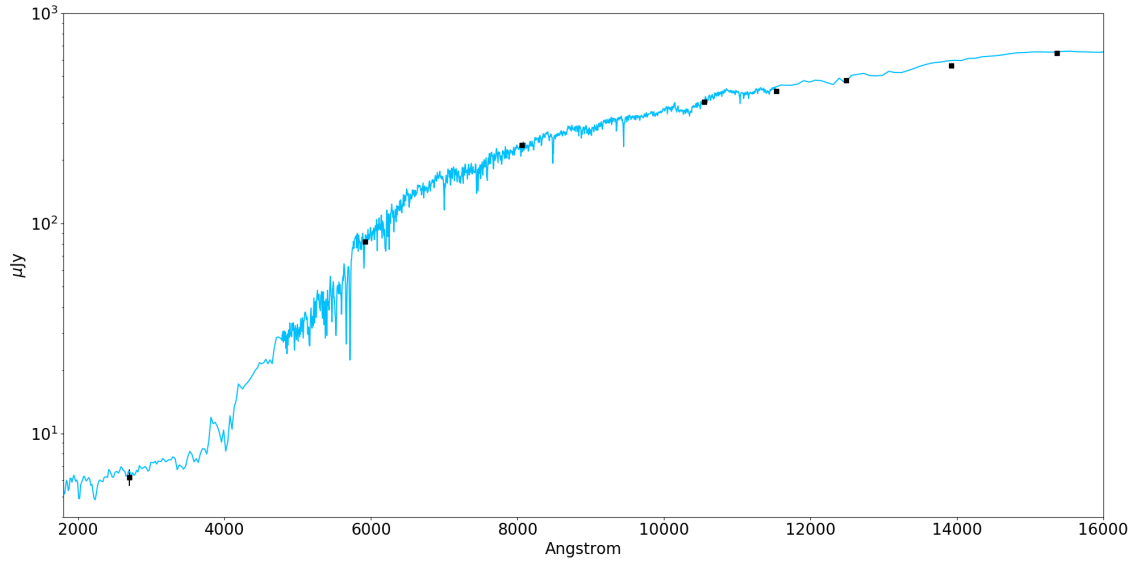


Figure 4.7: Best fit of cluster J1206's synthetic stellar population spectrum.

4.4.1 The central region of J1206

As mentioned earlier, we are particularly interested in the central region of J1206. We tested the performance of the nine models by comparing the predicted and observed images of the two resolved radial arcs, 4b and 7c, near the BCG. The model that best reproduces these two arcs would be, in principle, the best representation of the dark matter distribution in the center of J1206. For system 4, we used the counter-image 4c as a template for the source that is delensed and relensed by the nine lens models in order to predict the counter-image 4b. Counter-image 4a was not used due to its proximity to a cluster member. For system 7, we considered both counter-image 7a and 7d in the delensing and relensing process for reconstructing arc 7c. Comparing the predictions of arcs 4b and 7c with the data was not trivial in this case since we are making a comparison a posteriori. We did not look for models that minimize the residuals between the observed and predicted images. Instead, we have compared global properties (such as orientation, curvature, and position) of the predicted and observed arcs 4b and 7c after the models were derived using all available information. A classic residual analysis would be, in this case, not optimal since the models were derived not optimising the counter-images 4b and 7c, but optimising the predicted position of all arcs. The metrics used to compare the model with the data are a compromise between the standard metric, which uses only the position, and more complex figures of merit that could include additional parameters about the morphology of the arcs. We simply added an extra parameter (in addition to the standard position) to account for

the main morphological feature in each arc, radius of curvature in one case, and orientation in the other. A more complicated metric could, of course, be chosen but at the expense of introducing unnecessary correlations between the parameters.

First, we focused on the counter-image 4b (see Fig. 4.8), with a shape that can be well approximated by a straight line. From the original counter-image we fitted a straight line defined by the point where the arc intersects the BCG and the angle, α, between this line and the x-axis (left subfigure). We did the same process for the counter-images predicted using the different models (right subfigure). However, we were not able to select the first point (a) in the same way as we did with the original counter-image because there is no BCG covering the lower part of the predicted image. In order to tackle this we kept the same y-coordinate of the first point (a) from the original arc and only looked for the corresponding x-coordinate at which the predicted counter-image is maximum. From each one of these fits we computed two values: the x-coordinate of the lower point (a) and the angle (α) between the line of fit and the y-axis. Explanatory images of the steps followed are shown in Fig. 4.8. The predictions of counter-image 4b for all nine models are shown in Fig. 4.10.

For image 7c, we fitted it with a circle. Both from the original counter-image, and from the model-predicted images, we computed the coordinates of the two main knots ($7c_2$ and $7c_4$ according to our numbering criterion). We used these two coordinates to fit a circle to the whole arc in the real image and the predicted counter-images and we kept the coordinates of the center of these circles. From each one of these fits we saved three pairs of values: the coordinates of the knot $7c_2$, the coordinates of the knot $7c_4$ and the coordinates of the center of the fitted circle. Explanatory images of the steps followed are shown in Fig. 4.9. The predictions of counter-image 7c for all nine models are shown in Fig. 4.11.

We finally performed a Pearson's χ^2 test statistic in order to compare the performance of the different models in predicting counter-images 4b and 7c, and to find the models that work better. The values of the five parameters previously saved (two from 4b and three from 7c) are then used in the χ^2 test. To check for possible systematic effects, the previous process has been repeated several times under different conditions: using different counter-images as templates for the delensing and relensing process, centring the original images in different coordinates or using different filtered images to enhance the signal of the counter-images. Based on χ^2 values obtained and shown in Table 4.3, case 2b is the model which predicts with more accuracy both radial arcs. If we consider the two radial arcs separately, case 3b is the best model describing arc 4b, while case 2b is again the best model at reproducing 7c. The worst performers are case 1b for arc 4b, case 4b for

arc 7c, and case 4b again for the combination 4b+7c. Interestingly, both the best and worst performers correspond to the category of b-type models, that is models derived with the regular grid. In all cases, the arc 7c has a χ^2 that is significantly worse than arc 4b, indicating that there is still room for improvement in the model. This can be better appreciated in Fig. 4.9, where for the best model, case 2b, a small but clear offset between the predicted and observed image can still be seen.

Using case 2b, we were able to predict the position of counter-images that were not used in neither our analysis nor in Caminha et al. (2017). Our predictions for these counter-images are consistent with those in Caminha et al. (2017). These counter-images were not taken into account because they might be significantly deflected by massive and nearby early-type galaxies (22b, 23c, 24e according to the denomination of said paper) or because they are model predicted images in Caminha et al. (2017) but not observed in MUSE nor HST data (4d, 12d, 12e, 13e, 16d, 17c, 17d, 18c, 20c, 20d, 21e, 25c, 26c and 27c according to the denomination of said paper). Case 2b also predicts new counter-images for system 14 (RA = 12:06:15.9459, Dec = -8:48:09.909), for system 18 (RA = 12:06:11.8139, Dec = -8:48:46.595), and system 21 (RA = 12:06:11.2953, Dec = -8:48:40.112). These new predicted images are shown in Fig. 4.12.

Model	χ^2_{4b}	χ^2_{7c}	χ^2_{4b+7c}	M_1 (10^{12} $M_\odot$)	M_2 (10^{11} $M_\odot$)
1a	5.6	27.2	32.8	4.31	2.54
1b	12.1	21.2	33.3	3.35	8.02
2a	9.4	31.3	40.7	4.54	-
2b	4.6	14.1	18.7	5.27	-
3a	5.4	19.1	24.5	4.55	-
3b	1.4	24.6	26.1	7.37	-
4a	5.6	22.9	28.5	4.47	-
4b	6.2	47.1	53.3	7.57	-
5a	3.3	34.4	37.7	4.32	-

Table 4.3: χ^2 values and total mass of the galaxies in the compact component (M_1 for layer 1 and M_2 for layer 2, containing the BCG) for each one of the models considered to the reconstruction of counter-images 4b and 7c.

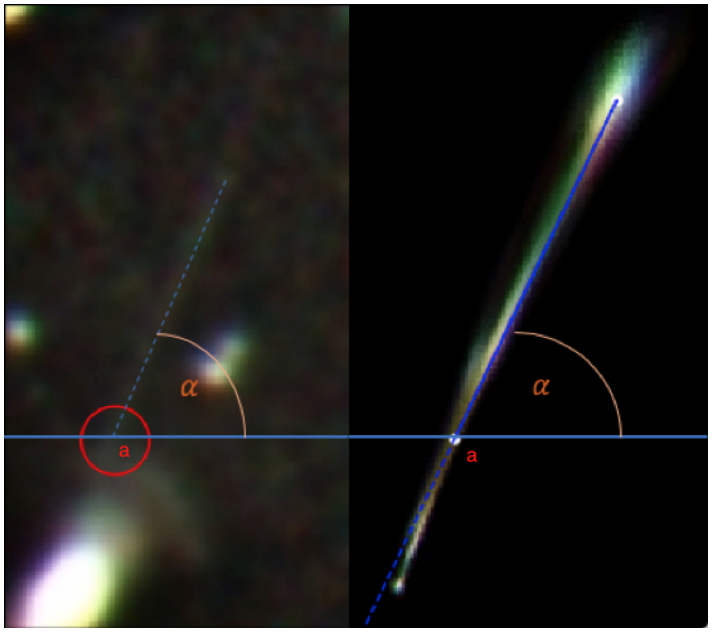


Figure 4.8: Images of the analysis performed with counter-image 4b. *Left*: Original counter-image 4b, where a blue dashed line has been used to approximate its shape, the position where this arc intersects the BCG is surrounded by a red circle (a), and the angle (α) between arc 4b and the x-axis is shown. *Right*: Counter-image 4b predicted with case 2b, where position (a) is fixed to the same y-coordinate selected in the *left* panel and x-coordinate is chosen to correspond with a maximum of the rebuilt image. This counter-image is fitted to a straight line, which has a certain angle (α) with respect to the x-axis. This process is repeated for every model considered.

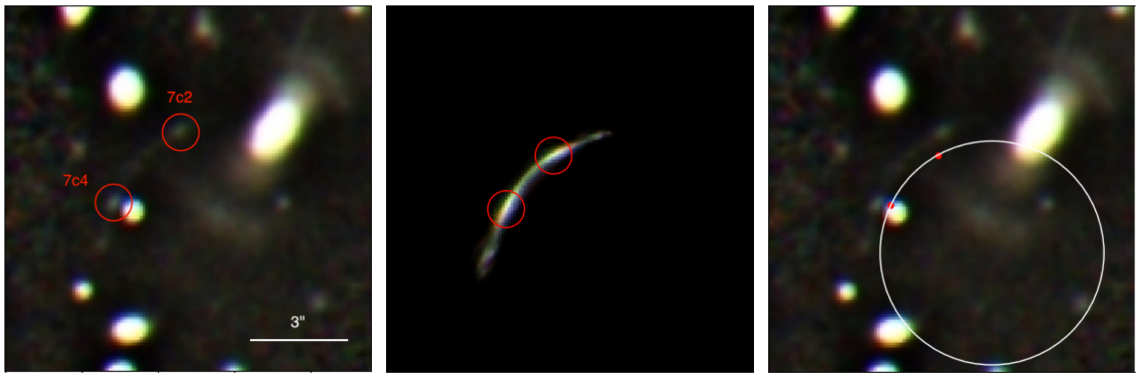


Figure 4.9: Images of the analysis performed with counter-image 7c. *Left*: Original counter-image 7c with its two knots $7c_2$ and $7c_4$ indicated with red circles. *Middle*: Counter-image 7c predicted with case 2b being its two knots highlighted. *Right*: Original counter-image 7c (*left* image) combined with the fit to a circle of the two knots (red dots) from the above prediction (*middle* image) so as the shift between both pairs of knots (original and predicted) can be better appreciated.

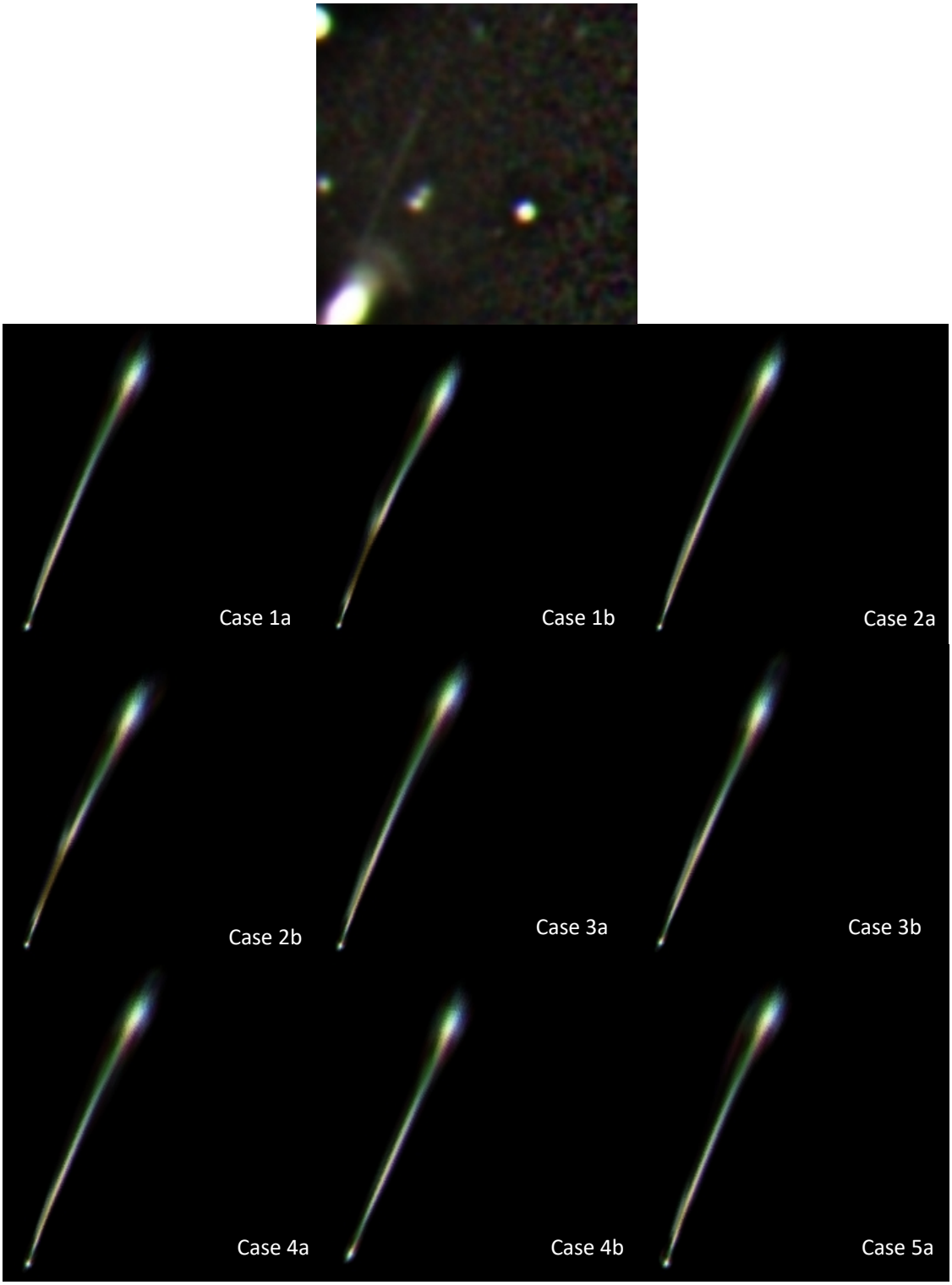


Figure 4.10: Original and predicted images of arc 4b, using the delensed counter-image 4c as a template of the source, for the nine different models discussed at the end of Sec. 4.3. The field of view and center of the images are the same in all panels. The image used was filtered to reduce the light contribution from member galaxies of the cluster.

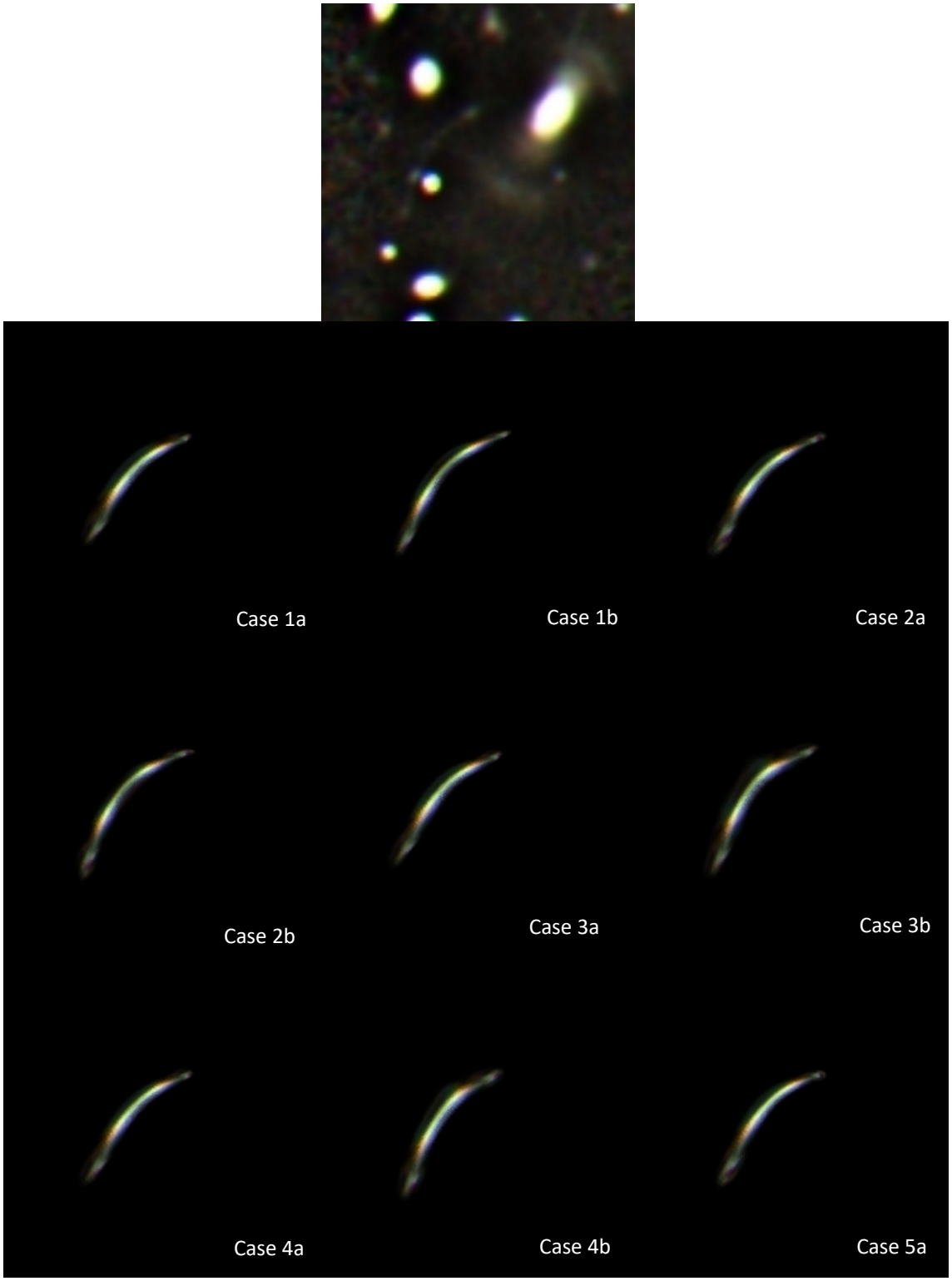


Figure 4.11: Original and predicted images of arc 7c, using the delensed counter-image 7a as a template of the source, for the nine different models discussed at the end of Sec. 4.3. The field of view and center of the images are the same in all panels. The image used was filtered to reduce the light contribution from member galaxies of the cluster.

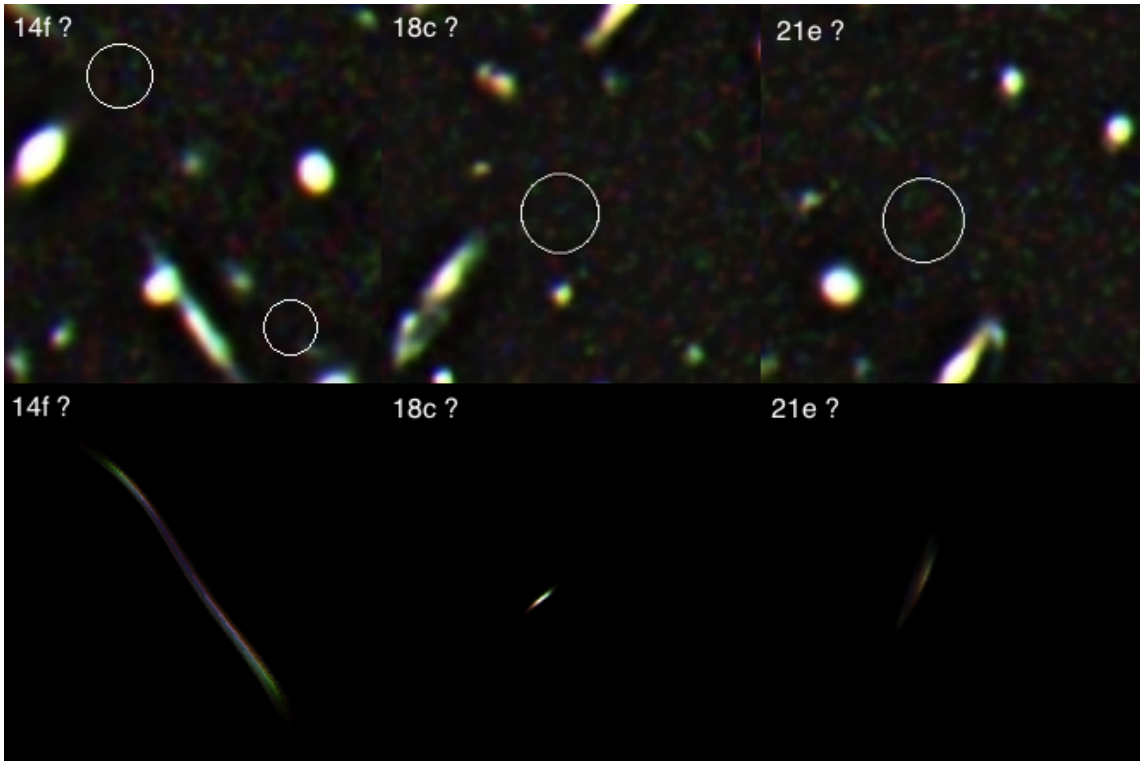


Figure 4.12: New counter-images predicted by case 2b for multiple-image systems 14 (left), 18 (middle) and 21 (right). The positions where these counter-images should be seen are indicated with white circles.

4.4.2 Mass distribution and mass profile

The two-dimensional mass distributions of the soft component (grid) for case 2b, our best-predict model, and case 3a, best-predict model using the multiresolution grid, are shown in Fig. 4.13, and they are overlaid with an image of the cluster. In Fig. 4.14, we show a comparison of the total projected mass profile of J1206 as estimated by galaxy kinematics (Biviano et al., 2013), by the strong lensing method performed in Caminha et al. (2017), and by the strong lensing analysis carried out in this work (Manjón-García et al., 2020). This projected mass profile is calculated by integrating the corresponding density profile along the distance. A very good agreement among the three profiles can be seen. The density profile of the galaxy cluster J1206 from case 2b is shown in Fig. 4.15. Both the projected total mass density profile (blue solid line) and the dark matter density profile (black dotted line) are shown.

In this analysis we have used the generalised NFW model shown in Eq. (3.46) to fit our estimation of the dark matter density profile within cluster J1206. After assessing the χ^2 values of several models to the measured dark matter density profile from case 2b, we found a reasonably good fit by using a gNFW density profile with $r_s = 167$ kpc (virial radius $r_{200} = 2$ Mpc and concentration parameter $c_{200} = 12$), $\rho_s = 6.7 \times 10^6\ M_\odot \mathrm{kpc}^{-3}$ and $\gamma_{gNFW} = 0.7$. This fit, performed within the range where strong lensing constraints are available, is shown in Fig. 4.15. These results differ with those obtained in previous works with this cluster. The gNFW parameters attained in Caminha et al. (2017) are $r_s = (300 \pm 3)$ kpc, $\rho_s = (1.91 \pm 0.3) \times 10^6\ M_\odot \mathrm{kpc}^{-3}$, and $\gamma_{gNFW} = 0.91 \pm 0.04$. The resulting gNFW density profile when using these values is also included in Fig. 4.15. We note that our profile at radii larger than 200 kpc falls below that of Caminha et al. (2017). Although this may be an artifact in our model due to memory effects of the initial condition (which are more severe in areas where the number of constraints is poor), it may also be a real feature as we are not assuming any specific profile for the halo at large radii. After comparing with the dynamical results from Biviano et al. (2013) shown in Caminha et al. (2017), the dynamical profile falls also faster than the profile in Caminha et al. (2017), and more in line with our profile at large radii. The best-fit NFW model scale radius, virial radius, and concentration parameter obtained in Biviano et al. (2013), by combining the results of two different kinematic methods, are $r_s = (0.35^{+0.14}_{-0.09})$ Mpc, $r_{200} = (1.96^{+0.14}_{-0.09})$ Mpc, and $c_{200} = (5.6 \pm 1.9)$. This last NFW model is also very close to the best-fit NFW model, with $r_s = (0.34 \pm 0.06)$ Mpc; $r_{200} = (1.96 \pm 0.11)$ Mpc; and $c_{200} = (5.8 \pm 1.1)$, obtained by the combined strong and weak lensing analysis published in Umetsu et al. (2012). Our estimate of the scale radius is, therefore, smaller than those of these previous

works.[4] Since the number of strong lensing constraints decreases the further we move away from the cluster center, and there are no constraints beyond 336 kpc, our method is not sensitive to the virial radius scale. Accordingly, the boundaries we achieve for the concentration parameter and the scale radius using our data are possibly not as robust as the ones provided by other data sets, involving for instance weak lensing.

The giant arc 4b is particularly interesting since its extreme elongation could in principle be explained by a very shallow profile in the case of a circularly symmetric lens model. This hypothesis is exciting since it could be the result of self-interactions in the dark matter fluid. Alternatively, the large elongation of arc 4b could be explained if the dark matter halo is oriented in the direction of the arc. This would result in a locally shallow potential in that direction, which is the requirement for long elongated radial arcs. When looking at the distribution of dark matter in the central region (see Fig. 4.13), the dark matter halo has similar morphology, and is oriented in the same direction, as the BCG. That is, the dark matter halo is oriented in the direction of the giant arc 4b. This fact suggests that one does not need to rely on a cross section for dark matter in order to explain the morphology of arc 4b. When fitting the value of the inner slope assuming a gNFW profile, we obtain a slope of $\gamma = 0.7$. This slope is smaller than the canonical $\gamma = 1$ predicted by standard CDM models but still relatively far to the expected shallow slope in self-interacting models with cross sections of the order of (or larger than) 1 $\mathrm{cm^2 g^{-1}}$, where one expects $\gamma \approx 0$ (Rocha et al., 2012).

Finally, we found an interesting result in the distribution of sources in the source plane. We estimated the positions of all 27 background galaxies using the lens model 2b. These galaxies are listed in Table 4.4, and shown in Fig. 4.16. Among these, the background galaxies with ID = 4, 5, 6, and 7 are all in a narrow redshift interval, $1.424 \leq z \leq 1.426$, and probably forming a group at this redshift since they cluster in the source plane both spatially and in redshift. A reconstructed version of the source plane with these four sources is shown in Fig. 4.17. We can see how galaxies 4 and 5 are already interacting, and that galaxies 6 and 7 are less than 10 kpc apart, thus in the early phase of an interaction or soon after a first encounter.

[4] We also performed a fit to a NFW density profile finding a good fit for $r_s = 178$ and $\rho_s = 4.56 \times 10^6$ $\mathrm{M_\odot kpc^{-3}}$. This estimation of the scale radius is smaller than in previous works too.

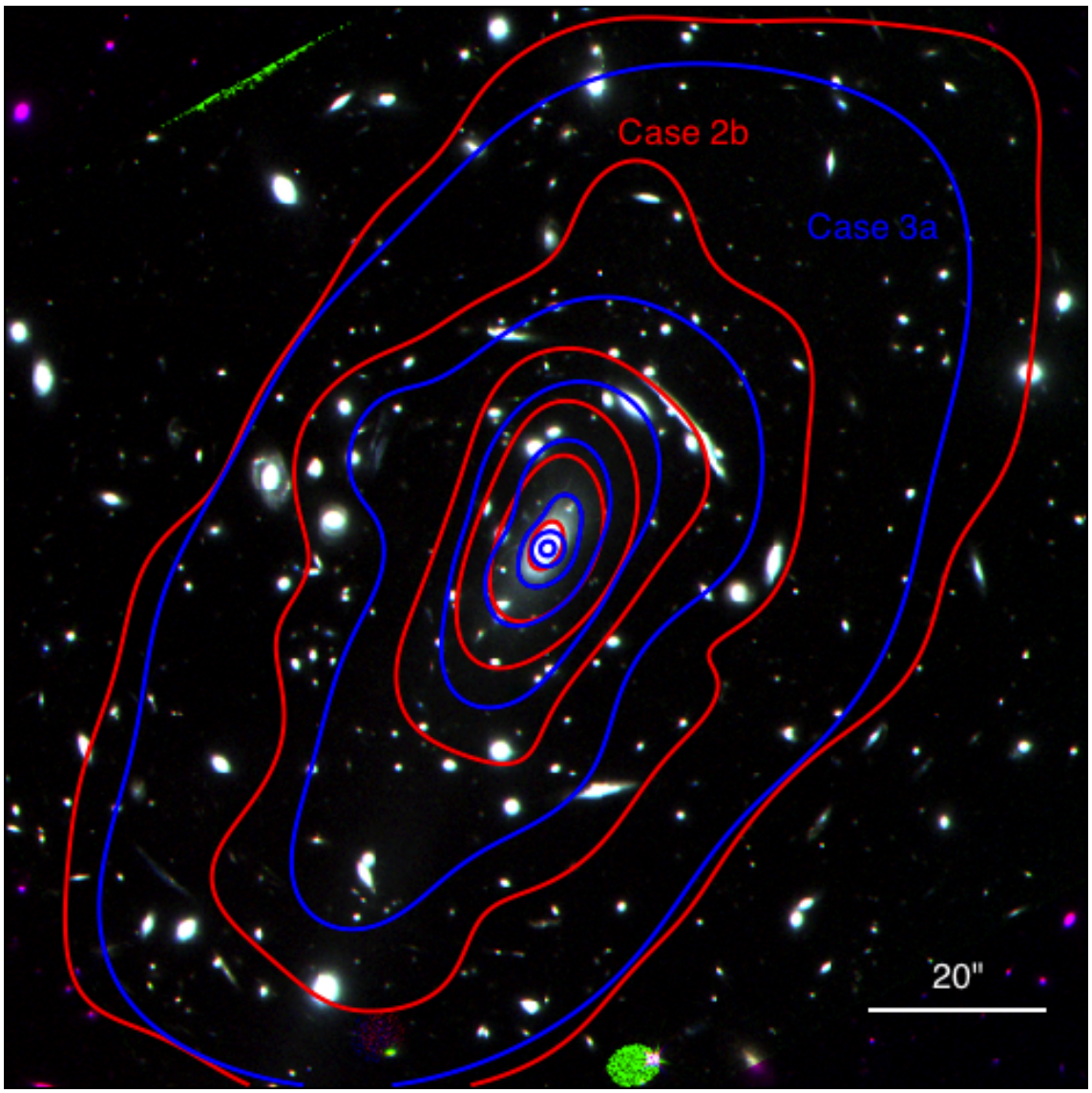


Figure 4.13: Contours of the soft component (grid) of the mass distribution for case 2b (in red) and case 3a (in blue) overlapped with an image of the cluster. The orientation of the image is not north-south, but rotated 44° counterclockwise.

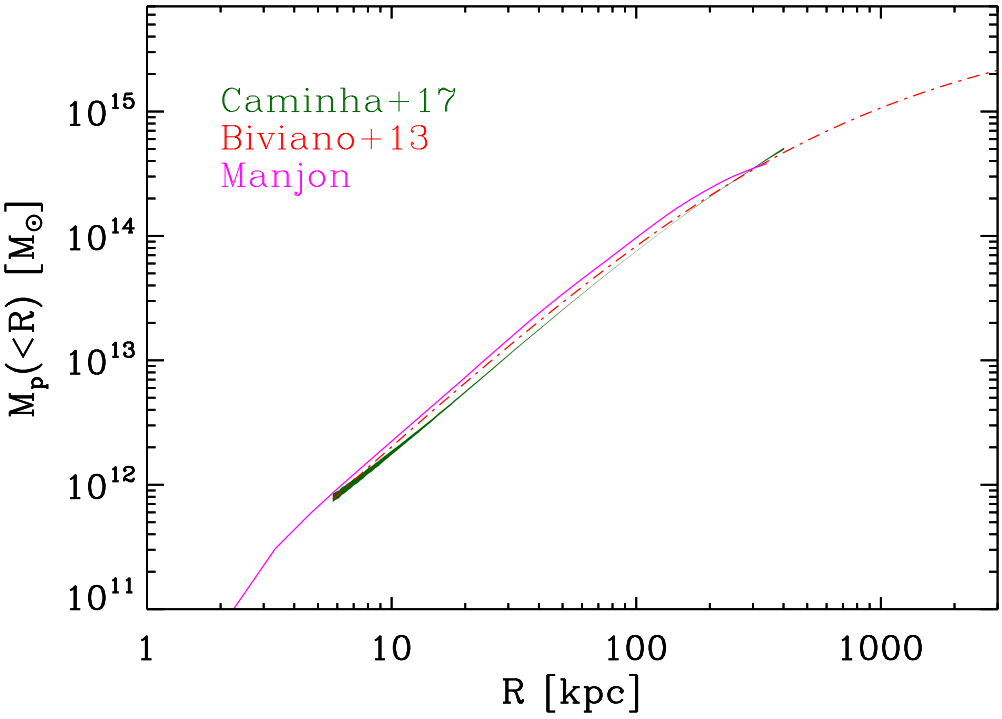


Figure 4.14: Comparison of the total projected mass profile of J1206 as estimated by galaxy kinematics in Biviano et al. 2013 (red), strong lensing in Caminha et al. 2017 (green) and strong lensing in this work (pink). Courtesy of Andrea Biviano.

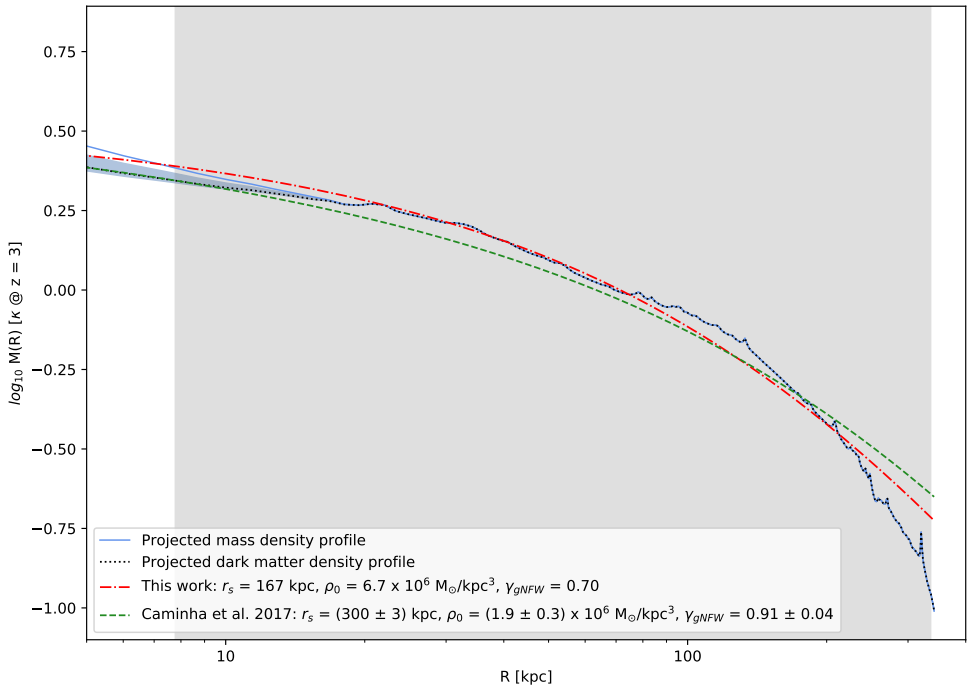


Figure 4.15: Mass profile in terms of the critical surface mass density (κ computed at z_s = 3) for the model that best reproduces the counter-images 4b and 7c (case 2b). The blue solid line is the mass density profile calculated with our method while the black dotted line is the dark matter mass density profile estimated after separating the stellar component of the BCG. The blue shaded region represents the 3σ uncertainty interval. The red dashed-dotted line is the best fit to a gNFW density profile achieved in this work while the green dashed line corresponds to a gNFW density profile considering the parameters attained in Caminha et al. (2017). The gray shaded area spreading from 7.8 kpc to 335.8 kpc (1.37" to 59.1") covers the region where all the multiple images considered in this work (and listed in Table 4.2) are placed.

N°	RA (deg)	σ_{RA} (arcsec)	Dec (deg)	σ_{Dec} (arcsec)
1	181.5502	0.1416	-8.8017	0.1105
2	181.5488	0.0684	-8.8005	0.0540
3	181.5501	0.0718	-8.8008	0.1029
4	181.5508	0.0222	-8.80132	0.2144
5	181.5509	0.0568	-8.80133	0.1017
6	181.5499	0.1128	-8.8021	0.0911
7	181.5500	0.2575	-8.8018	0.1000
8	181.5509	0.0846	-8.8016	0.0275
9	181.5488	0.0260	-8.8009	0.0426
10	181.5503	0.0648	-8.8025	0.0978
11	181.5553	0.0750	-8.8019	0.0728
12	181.5496	0.0318	-8.8025	0.0105
13	181.5515	0.1679	-8.7992	0.0839
14	181.5572	0.0289	-8.8025	0.0695
15	181.5518	0.0568	-8.8012	0.0392
16	181.5512	0.0998	-8.8012	0.1452
17	181.5511	0.0323	-8.8028	0.0076
18	181.5506	0.0105	-8.8031	0.0396
19	181.5548	0.0494	-8.8016	0.0166
20	181.5517	0.0651	-8.8006	0.0427
21	181.5496	0.0956	-8.8014	0.0651
22	181.5477	0.0076	-8.8004	0.0033
23	181.5552	0.0127	-8.8022	0.0868
24	181.5499	0.0972	-8.8020	0.1815
25	181.5527	0.0519	-8.8027	0.0407
26	181.5511	0.0865	-8.7980	0.1390
27	181.5493	0.0963	-8.8037	0.0730

Table 4.4: Positions (RA and Dec), estimated by case 2b, of the 27 background galaxies deflected by the cluster J1206, which have been spectroscopically identified. σ_{RA} and σ_{Dec} are the corresponding positional errors for the estimated positions.

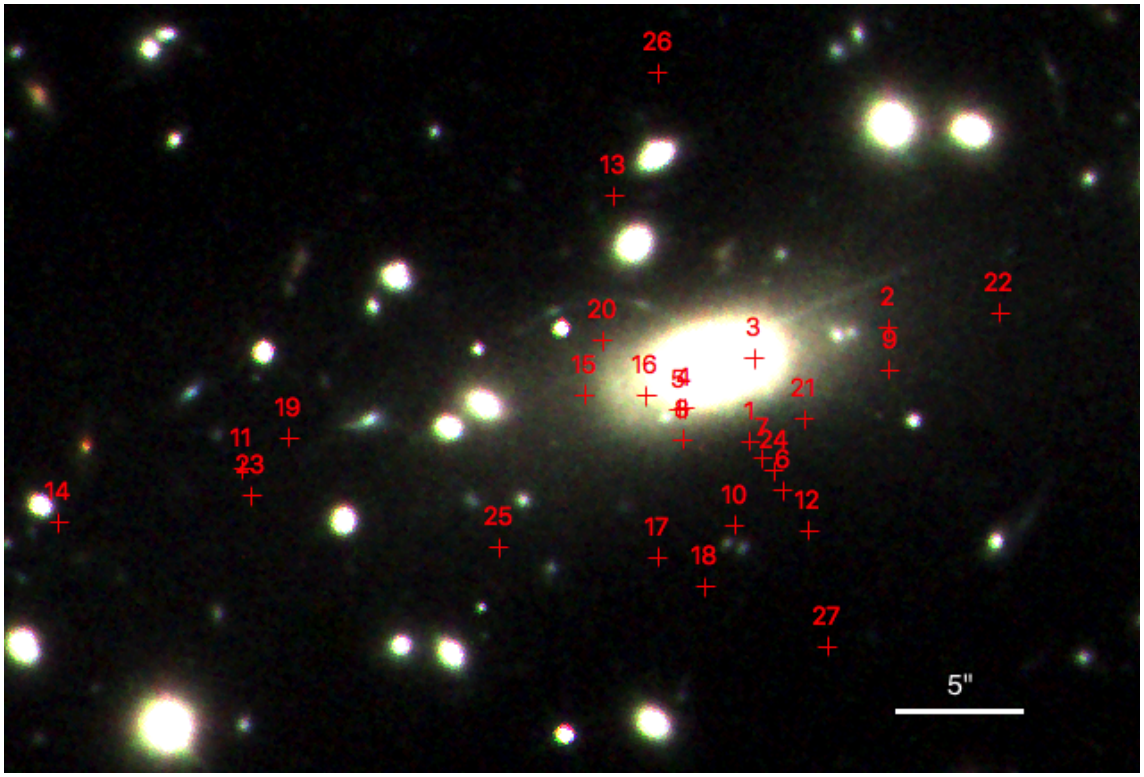


Figure 4.16: Image of cluster J1206 showing the positions of all 27 background sources estimated by case 2b. The field of view has 50 arcsec.

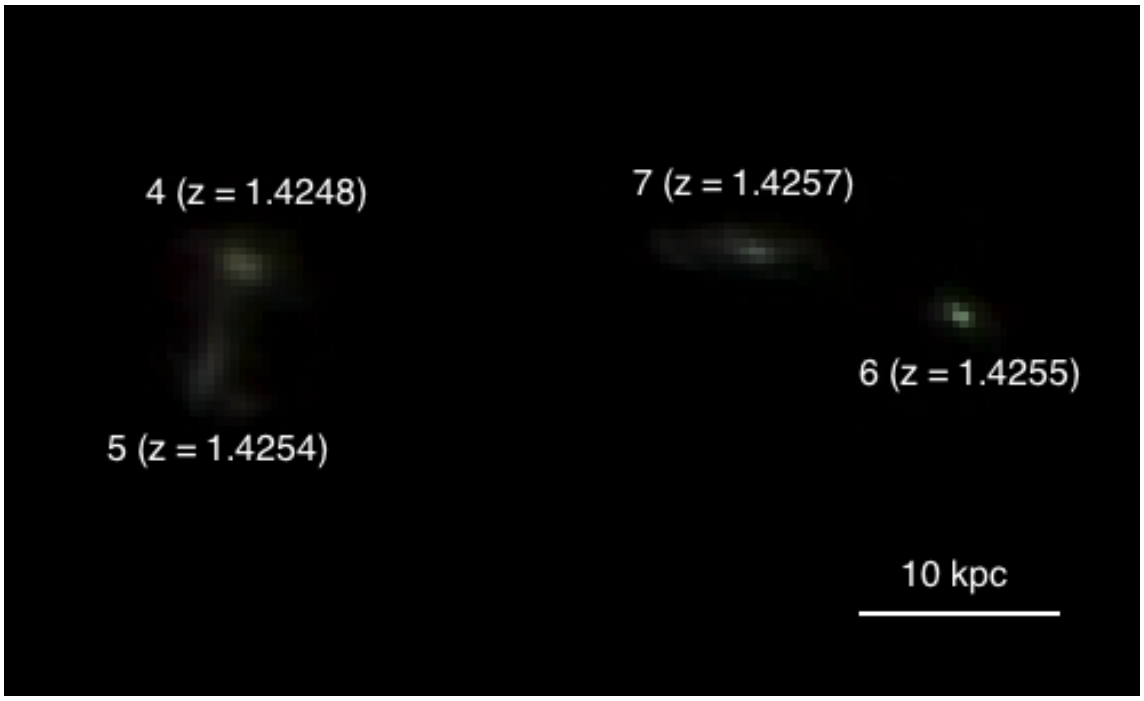


Figure 4.17: Reconstructed version of the source plane with background sources 4, 5, 6 and 7, which all have redshifts in the range [1.424-1.426].

4.5 Conclusions

In this chapter we have presented several free-form strong lensing models of the galaxy cluster MACS J1206.2-0847, at $z = 0.439$. We used public ACS and WFC3 Hubble data and the code WSLAP+ to derive the lens models. Our analysis was built upon spectroscopically confirmed multiple lensed sources identified using public deep MUSE observations in combination with CLASH-VLT imaging. The data set we used is based on the multiple image identifications from Caminha et al. (2017), but extended with the addition of further knots present in some well-resolved arcs. In total, 97 strong lensing constraints belonging to 27 background sources were used. The models considered had different configurations and were compared on the basis of how well they were able to reproduce two elongated radial arcs placed close to the BCG. Surprisingly, all models reproduced fairly well these two arcs, independently of the chosen configuration for the grid and choice for the mass-to-light ratio of the member galaxies. The geometry of the two giant radial arcs is mostly due to the elongated nature of the dark matter halo, which is reproduced by all models. We estimated, and later subtracted, the stellar baryonic contribution in order to estimate the remaining contribution from the dark matter component in the very central region. After comparing the predicted versus observed central radial images for the range of models, we identified the best model (2b) that reproduces the central potential. This best model is well described by a gNFW profile with $r_s = 167$ kpc and $\gamma_{gNFW} = 0.7$. Our estimate of the scale radius disagrees with previous results of the mass distribution within this cluster, although it should be noted that the lack of constraints at large radii impedes a robust estimation of this scale radius. This estimate leads to a concentration parameter $c_{200} \sim 12$, which falls in the range predicted for relaxed clusters. This supports the conclusion from Biviano et al. (2013) that the cluster is in a relaxed dynamical state. The results from our lens modeling are not indicative of a shallow profile in the central region of the cluster, as it would be expected for self-interacting dark matter models. On the contrary, we find a slope close to the expected slope for standard CDM models above 10 kpc. This is an important conclusion since our model makes no assumptions about the mass profile in the inner region (that could bias the constraints in the inner slope), other than assuming that the baryonic component traces the luminous mass.

Chapter 5

Neural networks theory

The analysis of strong gravitational lensing phenomena has proven to be an extremely useful cosmological tool. They can provide direct estimations of the dark matter distribution in the central regions of galaxies and galaxy clusters, and also be used to constrain cosmological parameters such as the Hubble constant. The only inconvenient is that strong lensing is a rare phenomenon, with only a few systems expected from surveying thousands of objects. The largest sample of strong lensed systems from a single survey until date, with more than 100 lenses observed, has been provided by the Sloan Lens ACS Survey (SLACS) (Bolton et al., 2008; Shu et al., 2017). Wide scale surveys, with millions of galaxy observations, are key to obtaining statistically significant samples of strong lenses. The advent of huge amount of data from the current and upcoming new generation of surveys, which will observe many more strongly lensed systems, could seem to solve this problem. The Euclid space telescope (Laureijs et al., 2011), due for launch in 2022, and the Large Synoptic Survey Telescope (LSST; LSST Science Collaborations, 2009) are foreseen to resolve $\sim 10^5$ strong gravitational lens systems (Refregier et al., 2010; Oguri and Marshall, 2010; Pawase et al., 2014; Collett, 2015). The Square Kilometer Array (SKA; Rawlings and Schilizzi, 2011) is also expected to observe an amount of radio-loud gravitational lenses of the same order of magnitude (McKean et al., 2015). Focusing now on ongoing surveys, we can point out optical wide surveys, such as the Kilo-Degree Survey (KiDS; de Jong et al., 2013) or the Dark Energy Survey (DES; The Dark Energy Survey Collaboration, 2005), which are expected to find the order of $\sim 10^3$ lenses (see e.g. Collett, 2015). Current infrared surveys also contribute to increase the statistics of detected lenses, as we discussed in Sec. 2.7 where a sample of 40 potential lens candidates in H-ATLAS were identified (Manjón-García et al., 2019). Sub-mm observations from *Herschel* (Negrello et al., 2010) and the South Pole Telescope (Carlstrom et al., 2011), together with deeper, high-resolution observations from the

Atacama Large Millimeter/sub-millimeter Array (ALMA; Wootten and Thompson, 2009) or the future Wide Field Infrared Survey Telescope (WFIRST; Akeson et al., 2019), are expected to provide several hundred of new lenses (González-Nuevo et al., 2012; Negrello et al., 2017).

However, traditional searches for lens features (i.e., arcs and rings) have exclusively relied on the visual inspection of the targets. And this is impractical for such volume of data. Therefore, the development of automatic lens finder techniques has awakened great interest recently. At first most of these techniques were based on the identification of arc-like and ring-like shapes (Lenzen et al., 2004; Horesh et al., 2005; Alard, 2006; Estrada et al., 2007; Cabanac et al., 2007; Seidel and Bartelmann, 2007; Kubo and Dell'Antonio, 2008; More et al., 2012). Nevertheless, these methods suffered from strong contamination and still required important eye-ball inspection of thousands of candidates per square degree. Maturi et al. (2014) followed the same approach but together with a color selection, which proved to be very effective in isolating the most promising arc candidates and reducing drastically the sample contamination. Other methods proposed consisted of subtracting the light of the central galaxies using multiband images, and then analysing the image residuals (Gavazzi et al., 2014; Joseph et al., 2014). The next natural step consisted of taking advantage of the recently popular machine learning techniques. In this sense, we can cite, for example, the identification of strong lenses using support vector machine (SVM) algorithms (Hartley et al., 2017). But the most successful supervised learning models in working with images are the Convolutional Neural Networks (CNNs). CNNs (Fukushima, 1980; LeCun et al., 1998, 2015) are a state of the art class of machine learning algorithm, particularly suitable for image recognition tasks since they are able to automatically extract high-level features at the pixel level. Even though they were theoretically developed in the 1980s and the 1990s, the hardware of the time could not meet their computational demands. Only recent innovations in hardware and the arrival of large labeled astronomical data sets has allow them to outperform other algorithms. The first application of CNNs to astronomical data sought to classify SDSS spectra (Hála, 2014). Subsequently, they have been used for the morphological classification of galaxies (Dieleman et al., 2015; Huertas-Company et al., 2015; Domínguez Sánchez et al., 2018), for measuring photometric redshifts (Hoyle, 2016; Pasquet et al., 2018), and for distinguishing between stars and galaxies (Kim and Brunner, 2017). And recently, they are intensely employed for the identification of strong lenses (Petrillo et al., 2017; Petrillo, 2019; Petrillo et al., 2019; Jacobs et al., 2017, 2019; Schaefer et al., 2018; Lanusse et al., 2018; Pourrahmani et al., 2018; Davies et al., 2019). Moreover, they are used for measuring the parameters of the lensing systems (Hezaveh et al., 2017), as well as their uncertainties (Perreault Levasseur et al., 2017), for infering the masses of galaxy clusters (Ho et al., 2019), and in the reconstruction of

undistorted images of background sources in strongly lensed systems (Morningstar et al., 2019).

Bearing all this in mind, we decided to follow a convolutional neural network approach to search for gravitational strong lenses in astronomical images. In this chapter we proceed to explain the main theory behind artificial neural networks, focusing on convolutional neural networks. We recommend reading Fei-Fei et al. (2020) for further details on this topic. Finally, we describe the network architecture we will use in the following chapters for the identification of strong lenses.

5.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) are computational models inspired by the behavior of the biological neural networks in the human brain. They consist of one or more layers of neurons, also called nodes or calculation units, linked by connections. These neurons are the basic unit of computation in a neural network. Each neuron in a layer receives one or more input values from the neurons placed in the previous layer, or from an external source if they belong to the *input layer*. Each one of these inputs has an associated weight (w), which is assigned depending on its relative importance to the other inputs. All neurons have an additional learnable parameter which acts as a bias offset (b). Neurons perform a weighted linear combination of all the received values and add the bias. Finally, a scalar function f, known as the *activation function*, is applied to this weighted sum in order to compute a scalar output y. The process described is shown in Fig. 5.1. In that outline, the neuron takes up to n different numerical inputs x_i, and associates them the same number of weights w_i. We will come back to talk about the activation function and bias after explaining the network architecture and the learning process.

The first and simpliest type of ANN designed was the feedforward neural network. It contains multiple neurons arranged in several layers. There are no cycles or loops in this kind of network. The information moves only forward from the input layer to the output layer, going through the hidden layers, if there were any. An example of a feedforward neural network is shown in Fig. 5.2. As can be seen in it, nodes from adjacent layers can be connected to each other, and have their corresponding weights associated.

The architecture of the network is built upon three types of nodes. The input nodes are the neurons that provide to the network information from the outside, and are exclusively arranged in the input layer. They do not perform any computation, just pass on the information to the next layers. The output nodes are the neurons that perform the last

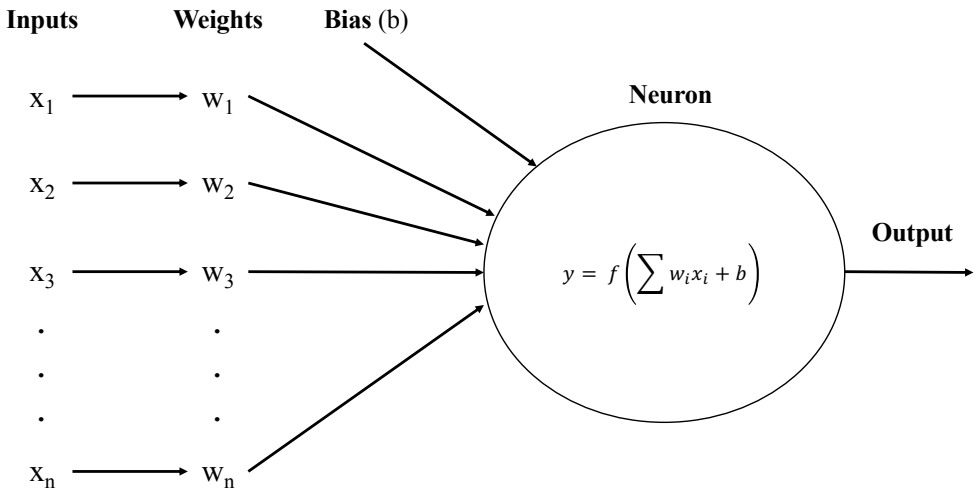


Figure 5.1: Scheme of the operation of a single artificial neuron.

computations and provide us with the final result of the network estimation. They are organized forming a single output layer. Lastly but not least, the hidden nodes are all the neurons placed between the input and the output layer. They are so called because they are not in contact with the outside world. They perform the main computations and convey the information from the input nodes to the output nodes. Hidden nodes are organized in hidden layers. Although a feedforward network only has one input layer and a single output layer, it can have from zero to multiple hidden layers. Feedforward neural networks can be split into two categories:

- **Single layer perceptron**: The simpliest feedforward network, which does not contain any hidden layer.
- **Multi layer perceptron**: This is the name given to any feedforward network that has at least one hidden layer.

5.1.1 Multi layer perceptron

A Multi Layer Perceptron (MLP) is a feedforward neural network that has one or more hidden layers. Having hidden layers allows this kind of ANN to learn both linear and non-linear functions, in contrast with what happens in the single layer perceptron, which only can learn linear functions. The ANN shown in Fig. 5.2 is an example of MLP with a

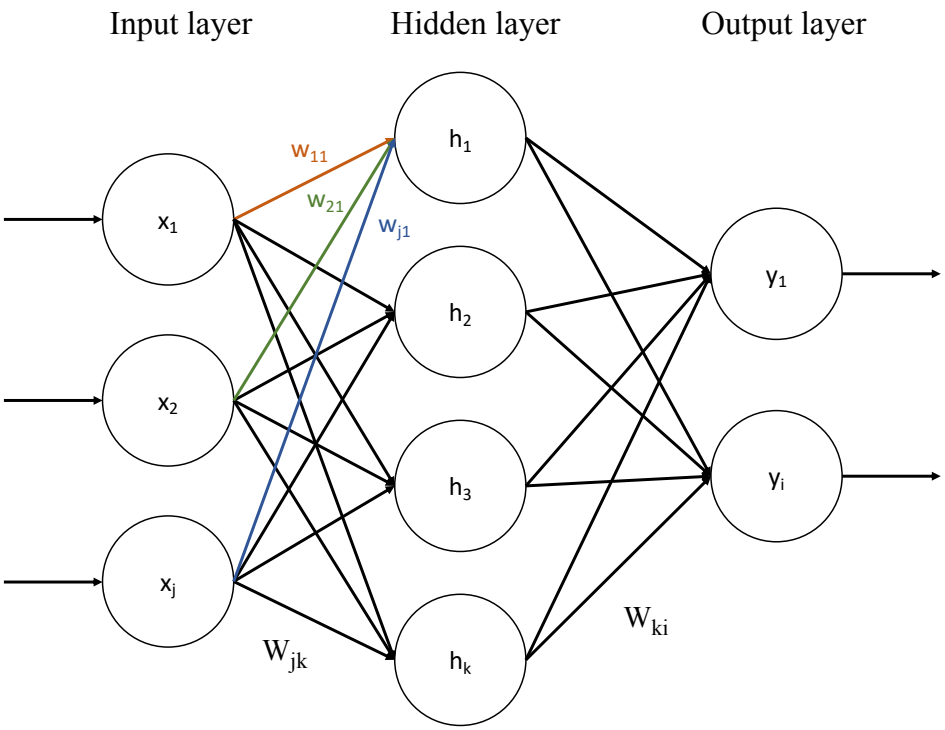


Figure 5.2: Example of a feedforward neural network known as multilayer perceptron. It consists of at least three layers of nodes: input, hidden and output layers.

single hidden layer. Although the input, hidden and output layers have been drawn with three, four, and two neurons, respectively, the last node in each layer has been labeled with the index j, k, or i, respectively, indicating that the number of neurons in these layers is arbitrary. In this sense, all connections will have weights associated with them, but only three weights (w_{11}, w_{21} and w_{j1}) belonging to three connections between the input and the hidden layer are shown in the figure. According to the neuron bevavior explained above, and summarized in Fig. 5.1, the output from any hidden neuron in the single hidden layer MLP shown in Fig. 5.2 is given by:

$$h_k = f\left(\sum_j w_{jk} x_j + b_k\right) \tag{5.1}$$

where w_{jk} are the weights associated to the connections between the chosen hidden neuron and the input layer, x_j are the different inputs coming from the input layer, and b_k is the bias corresponding to the hidden neuron considered. Following the only possible direction,

all the information from the hidden nodes arrives to the output layer. The resulting estimation provided by any output neuron in the MLP shown in Fig. 5.2 is given by:

$$y_i = f\left(\sum_k w_{ki} h_k + b_i\right) = f\left(\sum_k w_{ki} f\left(\sum_j w_{jk} x_j + b_k\right) + b_i\right) \tag{5.2}$$

where w_{ki} are the weights associated to the connections between the chosen output neuron and the hidden layer, b_i is the bias corresponding to the output neuron considered, and f is the activation function, which is assumed to be the same for all neurons in the MLP.

5.1.2 Backward propagation

One of the main properties of ANNs is their ability to learn from data, for either classification or regression. The different learning methods can be broadly classified into two categories based on the type of information fed into the network as training data. In *supervised learning*, training data are labeled, i.e., they consist of several input values (a_p) paired with the desired output (b_p) from the network for each of those input values, with $p = 1, ..., r$. On the other hand, when data are presented to the network without additional information and the network has to discover on its own the hidden patterns and categories, we are talking about *unsupervised learning*. From now on we will focus solely on supervised learning.

The input data set used to train the network is characterized by the features and the number of examples given. The number of features is equal to the number of nodes in the input layer of the network, which ranges within $j = 1, ..., n$. For instance, if an ANN was being trained to classify among different species of mammals, the features could be the height, weight or, hair length. This network would have an input layer with three nodes. On the other hand, an example consists of the values for each of the features, along with a truth label indicating the desired output of the network for this example. As it was earlier said, the number of examples ranges within $p = 1, ..., r$. In the mammals classifier, the examples would be, for instance, the label 'lion' with their corresponding feature values, the label 'koala' with their values, another label 'lion' with feature values slightly different from those in the first 'lion', etc. The training data set can have as many features and examples as available or as we want. In fact, it is partially true that the more examples and features the network sees, the better it can learn.

The number of output nodes in the output layer depends on the specific purpose for

which the network is being trained. In a binary classification problem between two groups or options, the output layer would have two nodes. One of them would output the probability of belonging to the first group while the other node would output the probability for the other option. Whether the examples considered are classified into one or the other group will depend on which of the two nodes scores a higher probability. This interpretation remains the same for multiclass classification problems, where there will be as many output neurons as classes. Going back to the classification of mammals, this network would have as many output neurons as the mammals we want the network be able to distinguish. The number of neurons in the output layer ranges within $i = 1, ..., m$.

The learning process of a MLP is performed by tuning the weights associated to all the connections in the network. In the supervised learning, the weights are usually obtained by minimizing some error function, conventionally called *cost* or *loss function*, that measures the difference between the desired output values and those calculated by the neural network. This training scheme is called *backward propagation of errors*. In first place, initial weights are set. They can be chosen randomly or assigned based on prior knowledge. Training data start going through the network and an output result is obtained for each input. Once the ANN has seen all the data for the first time, the loss function can be evaluated. There are several loss functions used in neural networks, but for now we are only defining one of them as an example. A fairly well-known example of loss function consists of calculating the sum of squared residuals for every individual input data:

$$L \equiv E(w) = \frac{1}{2} \sum_{i,p} (b_{pi} - y_{pi})^2 \tag{5.3}$$

where b_{pi} are the desired output values and y_{pi} are the output values predicted by the network, according to Eq. (5.2), in each one of the $i = 1, ..., m$ output nodes for the $p = 1, ..., r$ examples in the training data set. The errors calculated with the loss function at the output nodes can be then propagated back through the network. The stochastic gradient descent algorithm is then used to adjust all weights in the network with the aim of reducing the errors in the output layer. The change in the weights required so that we move down in the direction of steepest descent of the loss function is given by the following partial derivatives:

$$\begin{aligned} w'_{jk} &= w_{jk} - \eta \frac{\partial E(w)}{\partial w_{jk}} \\ w'_{ki} &= w_{ki} - \eta \frac{\partial E(w)}{\partial w_{ki}} \end{aligned} \tag{5.4}$$

where the apostrophe $'$ stands for the new weights, the absence of apostrophe stands for the old weights, and η is the learning rate or learning parameter. The learning rate is a decisive configurable hyperparameter that, every time the weights are updated, adjusts how much the network model is changed as a function of the estimated error. Choosing the learning rate may be tricky. A too small value could result in a long training process that could even get stuck, whereas with a value too large the training could be unstable and the weights achieved not optimal. It usually has a small positive value in the range between 0 and 1.

At this point, the output of the neural network and the loss function can be calculated again with the new weights. The network should perform better now than before since the weights have been adjusted to decrease the error in prediction. This is an iterative process, in which we will keep updating the weights until finding the ones that minimize the loss function, i.e., until the output error is below a predetermined threshold. Loss functions used in neural networks are non-convex functions, i.e., they may have several local minima that are not as deep as the global minimum we are interested in finding. Therefore, it can be hard to tell if the iterative process got stuck in a local minimum instead of reaching the global one. Once a successful learning is accomplished with the training data, the neural network is ready to work with new data.

To sum up, the learning problem in ANNs comes down to getting the right weights (w) for approximating a given set of input-output patterns $(a_{p1}, ..., a_{pn};\ b_{p1}, ..., b_{pm})$, in such a way that the output values $(y_{p1}, ..., y_{pm})$ from the network for certain input values $(a_{p1}, ..., a_{pn})$ are as close as possible to the desired output values $(b_{p1}, ..., b_{pm})$. Once these weights are reached, we have a trained ANN, which has learned from the labeled data and the error propagation, ready to work with new input data.

5.1.3 Activation function

The aforementioned activation function f is a non-linear function whose purpose is to introduce non-linearity into the output of the neuron. This is necessary since most of the data in the real world are non-linear. Activation functions receive the data and perform a certain fixed mathematical operation with them. There is a number of activation functions that may be encountered in practice, but the following are the most commonly used:

- Sigmoid: This function takes real-valued inputs and squashes them to range between 0 and 1 according to the following mathematical expression:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{5.5}$$

The behavior of this function can be seen in the left plot of Fig. 5.3. It turns large negative numbers into 0 and large positive numbers into 1. The sigmoid non-linearity has often been historically used given its straightforward interpretation of the firing rate of a neuron: from a full saturating response (1) to the absence of it (0). However, it is rarely ever used nowadays because using it involves three main drawbacks. The first disadvantage is that when the activation saturates at either 0 or 1, the gradient is almost zero at these regions. And if this local gradient is very small, the whole gradient will be killed in the backpropagation process. As a result, almost no signal will be left to flow through the neurons. This is known as the *vanishing gradient problem*. The second undesired effect is that if the initial weights are too large, most neurons will saturate and the network learning will be very poor. For this reason, caution must be taken when initializing the weights of sigmoid neurons. The last inconvenience is due to the fact that the outputs of sigmoid neurons are not zero-centered. If a neuron always receive positive data, the resulting gradient on the weights w will become either all positive, or all negative, depending on the gradient of the expression of the activation function f. This could introduce an undesired zig-zagging pattern in the gradient update. However, in the end this effect is somewhat mitigated because once these gradients are added up across many data, the update for the weights can end having variable signs.

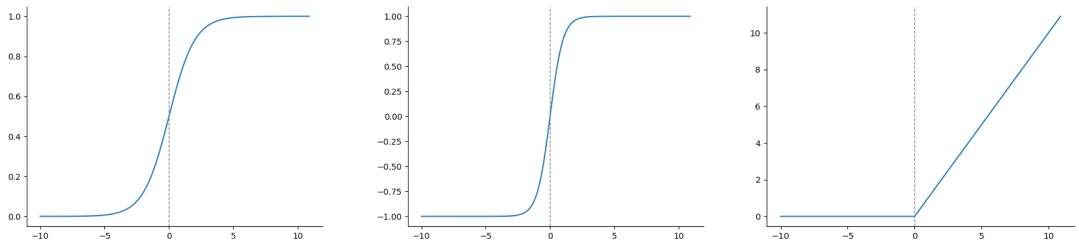


Figure 5.3: Sigmoid, tanh, and ReLU activation functions.

- Tanh: The hyperbolic tangent function receives a real-valued input and squashes it to the range [-1,1] according to the mathematical formula:

$$tanh(x) = 2\sigma(2x) - 1 = \frac{2}{2 + 2e^{-2x}} - 1 \tag{5.6}$$

As it can be noted in this equation, a tanh neuron is simply a scaled sigmoid neuron. This modification allows us to get neurons which continue saturating but whose outputs are now centered at zero. Therefore, in practice the tanh non-linearity is a better choice than the sigmoid one. Tanh function is affected by the vanishing gradient problem too. Due to the nature of the backpropagation optimization, gradients tend to get smaller and smaller as keep moving backwards. It implies that neurons in first layers learn rather slower than neurons in the last layers. This results in a decrease of the prediction accuracy of the network and a long training process. The behavior of this function can be seen in the middle plot of Fig. 5.3.

- ReLU: The Rectified Linear Unit (ReLU) function takes a real-valued input and thresholds it at zero, according to:

$$f(x) = \max(0, x) \tag{5.7}$$

Negative values are replaced with zeros while positive values keep their input value. The behavior of this non-linearity can be seen in the right plot of Fig. 5.3. This activation function was introduced for the first time to a dynamical network by Hahnloser et al. (2000). It has been proved by Glorot et al. (2011) and Krizhevsky et al. (2012) that ReLUs speed up a lot the convergence of the stochastic gradient descent compared to the previous mentioned functions. This is argued to happen because of its linear non-saturating form at the positive numbers range. In comparison to the sigmoid and tanh neurons, which involve time-consuming operations, ReLUs can be easily realized by thresholding a matrix of activations at zero. ReLUs are also not affected by the vanishing gradient problem. Another major improvement is that they are capable of giving a true zero value as a result, unlike tanh and sigmoid functions that only learn to approximate a zero output. This means that negative inputs can activate the hidden layers in the network to calculate true zero values. This property, called sparse representation, is really rewarding as it can accelerate learning and simplify the derived model.

ReLUs have become very popular in the last few years due to these advantageous features. Unfortunately, they are not exent of some flaws. The flow of a large gradient through these neurons might result in such a weights update that no data would ever be able to activate them. If this happens, the gradient coming from this unit will forever be zero from that point on, i.e., the ReLUs can irreversibly die during training. This is known as the "dying ReLU" problem. It is, therefore, important to have as few dead neurons as possible in the network. A proper setting of the learning rate η

can ensure this is less of a problem. If the learning rate is too high the weights may change to a value that causes the neuron to not get updated with any data again.

- Leaky ReLU: Leaky ReLUs arise as an attempt to sort out the problem of ReLUs' death. Instead of turning all negative values to zero, this function renders a small negative slope (of ≈ 0.01, or so) when $x < 0$:

$$f(x) = \begin{cases} \alpha x & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$$

where α is a small constant. The behavior of this function for $\alpha = 0.5$ can be seen in the left plot of Fig. 5.4. The results with this non-linearity are not always consistent. The slope in the negative zone can also be turned into a specific parameter of each neuron, as seen in He et al. (2015). The function in this case is known as Parametric Rectified Linear Unit (PReLU).

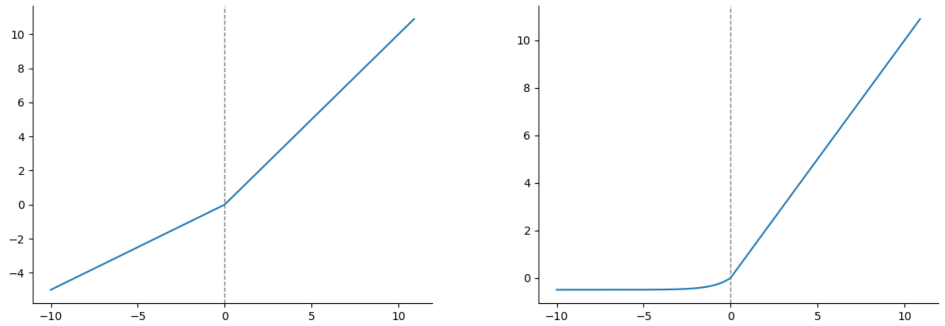


Figure 5.4: Leaky ReLU and ELU activation functions. A constant value higher than usual ($\alpha = 0.5$) was taken so that the behavior of each function when $x < 0$ can be well appreciated.

- ELU: The Exponential Linear Unit (ELU) is another attempt to fix the fragility problem of ReLUs. In this case, the linear behavior when $x < 0$ is replaced by an exponential behavior according to the following:

$$f(x) = \begin{cases} \alpha(e^x - 1) & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$$

This function is depicted, considering $\alpha = 0.5$, in the right plot of Fig. 5.4.

- Maxout: All types of neurons seen so far consist of applying a non-linearity on the dot product between the data and the weights: $f(w^t x + b)$. The maxout activation function is a different choice which was introduced by Goodfellow et al. (2013). This neuron is defined as follows:

$$f(x) = \max(w_1^t x + b_1, w_2^t + b_2) \tag{5.8}$$

  A maxout unit is a piecewise linear function that returns the maximum value among the n linear functions considered ($n = 2$ for this case). It is an activation function trained by our model. It can be observed that both the ReLU and the Leaky ReLU are special cases of this form. For example, for the ReLU we would have $w_1, b_1 = 0$. Therefore, the maxout neurons are not at risk of dying during the process, and offer a linear regime of operation without saturation. However, the total number of parameters to be trained by the network is doubled. This activation function was designed to be used along with the dropout regularization technique, about which we will talk in the next section.

- Softmax: The softmax function is an activation function that receives a vector of real-valued numbers, also called digits, and turns them into probabilities that sum to one, according to the following expression:

$$S(\boldsymbol{y})_i = \frac{e^{y_i}}{\sum_j e^{y_j}} \tag{5.9}$$

  In words, an exponential function is applied to each digit y_i of the input vector $\boldsymbol{y}$ and then these values are normalized by dividing by the sum of all these exponentials. The normalization ensures the sum of all components in the output vector $S(\boldsymbol{y})_i$ is one. This output vector represents the probability distribution of a list of potential outcomes. We mean that, before the softmax function is applied, there was no restriction for the vector values. But after applying it, every component is within the interval [0,1], and their sum will be 1, so that they can be interpreted as probabilities. For this reason, softmax is typically used as the activation function in the neurons of the output layer, to turn the non-normalized output of the network into a probability distribution over the predicted output classes.

It should not be forgotten to say that the importance of the aforementioned bias b lies in providing every node with a trainable constant value, apart from the normal inputs, which allows shifting the activation function to the right or left. This small room for maneuver may be critical for achieving a successful learning.

All things considered, the question arises of which of these activation functions should be used. The answer is rather straighforward. Softmax, hyperbolic tangent or linear activation are the best choices for the output layer. However, the best decision for the hidden layers of the neural network is to start using the ReLU non-linearity. As it was demonstrated by Glorot et al. (2011) and Krizhevsky et al. (2012), this activation function enables a better training of deeper networks compared to the sigmoid and tanh functions. It is currently the most widely-used and successful activation function. Although it is important to be careful with the learning rate chosen and watch out the fraction of dead neurons in the whole network. Leaky ReLU, PReLU, ELU or maxout are good alternatives to try if any persistent problem arise. Due to the vanishing gradient problem, it is not recommended to use the sigmoid and tanh functions in networks with many layers.

5.1.4 Regularization techniques

We have already made it clear that neurons are the essential elements in a neural network, and that they are arranged in three different types of layers depending on their purpose. However, what architecture should a neural network have to face a practical problem? How do we decide it? A ground fact to be noted first is that the deeper the neural network, the greater its predisposition to learn will be. As the size and the number of layers in the ANN are increased, more and more neurons can work together, and the space of learnable functions grows. ANNs with more neurons are able to express more complicated functions, i.e., they have greater capacity. It is sad to say, but this blessing comes along with its curse. Despite the fact that large neural networks can learn to classify more complicated data, they can also acquire such precise knowledge of the training data set that they are not able to perform well with an unseen data sample. This undesired consequence is known as *overfitting*. Overfitting occurs when a high capacity model fits the noise in the data instead of the underlying relationships of interest. This is precisely caused because the model is overly complex, with too many parameters. An overfit model can easily be diagnosed by monitoring its performance during training on both a training data set and on a validation data set. The validation data set is not used to adjust the weights on the neural network, but to verify that any increase in accuracy over the training data yields an increase in accuracy over data that the network has not seen earlier. If the accuracy over the training set increases, but the accuracy over the validation set decreases or remains equal, then the network is overfitting the training data. There are two ways to reduce overfitting: increasing the number of examples in the training data set or reducing the complexity of the network.

The capacity of a neural network model is defined by its architecture in terms of nodes and layers, and its weights. The complexity of an ANN can, therefore, be reduced by changing its structure (number of weights) or its parameters (values of weights). Based on this, it could seem preferable to use smaller neural networks, with fewer layers and fewer neurons per layer, if data are not large or complex enough. This thought is wrong. In practice, it is always better to use other existing methods to control overfitting rather than decreasing the number of neurons. This is why it is more difficult to train small neural networks with local methods, such as the gradient descent. Although their loss functions have few local minima, many of them have high loss and are easier to converge to. On the contrary, larger neural networks contain significantly more local minima, but they turn out to be much better reducing the loss. The work of Choromanska et al. (2015) studied the behavior of the loss surfaces in feedforward neural networks and their results pointed in this direction. The probability of finding a bad local minimum is not zero for a small network, and it decreases quickly as the network size gets larger. Most local minima in large networks are essentially equivalent from the point of view of the error decrease reached, and rely less on lucky random initializations. That is why there is no point in struggling to find the global minimum in the training, as it can even lead to overfitting.

Larger neural networks are proven to perform better than smaller networks, but their higher capacity must be suitably addressed in order to control overfitting. The procedure of limiting the capacity of a model by making slight modifications to the learning algorithm is called regularization. Regularization is used so that the model trained by the ANN can generalize properly to new data, i.e., make accurate predictions for data it has not been trained with. Some of the most known regularization techniques to prevent overfitting from occurring will be introduced below:

- Early Stopping: As it was said earlier, a validation data set can be used to detect when overfitting starts during the supervised training of a neural network. The technique of stopping training at the point when the performance on the validation sample starts to downgrade is called early stopping (Prechelt, 1998). This point has to be carefully evaluated since insufficient training will lead to an underfit model that performs poorly on the training data. The exact moment to stop training is when a decrease in the error estimation on the training data does not yields an error decrease in the validation data, but it rather remains the same or even could increase (see Fig. 5.5). This is a simple and widely used approach to tackle overfitting, which can be used either alone or in conjunction with other regularization techniques.
- L2 Regularization: This and the following technique are aimed to penalize the model during training based on the magnitude of the neuron activations. They do so by

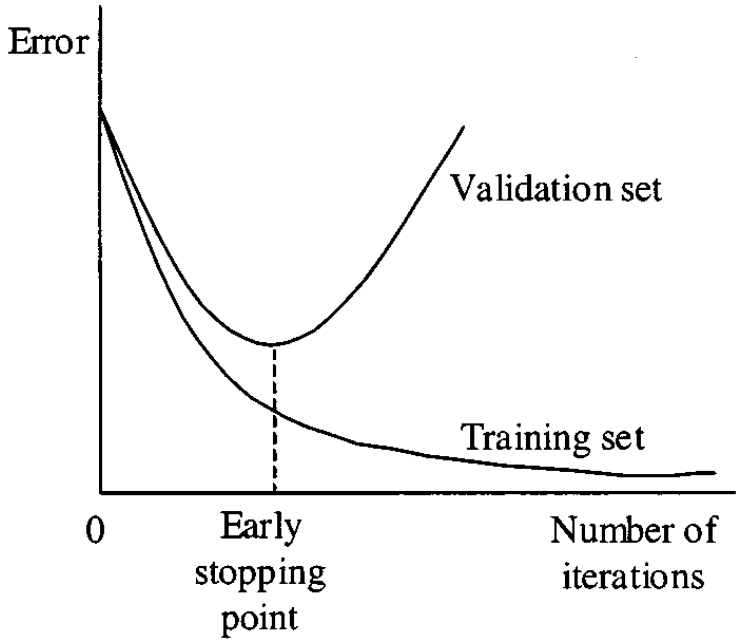


Figure 5.5: Example of the evolution of the learning for a training and a validation samples, showing the early stopping point. Plot taken from Gençay (2001).

adding penalty terms to the loss function. The L2 penalty seeks to minimize the squared magnitude of the weights according to the following:

$$L(w_{ij}) \equiv L(w_{ij}) + \frac{1}{2}\lambda \sum_{i,j} w_{ij}^2 = \sum_{i,p} (b_{pi} - y_{pi})^2 + \frac{1}{2}\lambda \sum_{i,j} w_{ij}^2 \tag{5.10}$$

where the MLP network from Fig. 5.2 and the squared residuals loss function from Eq. (5.3) were assumed. In this equation, w_{ij} stands for the matrix of the weights (w_{jk}, w_{ki}) involved since all of them are updated in the backpropagation algorithm, and λ is the regularization strength. Turning it into words, a term $\frac{1}{2}\lambda w^2$ is added to every weight in the network. The factor $\frac{1}{2}$ is used so the resulting gradient of the term with respect to the weight w is just λw instead of $2\lambda w$. In fact, during the parameter update every weight decays linearly towards zero according to $w_{ij} = w_{ij} - \lambda * w_{ij}$. The intuitive interpretation of this regularization is that extreme weight vectors are heavily penalized, while diffuse weight vectors are preferred. This has the effect of pushing the network into having all of its inputs contributing a little rather than having a few inputs contributing a lot. This is one of the most common forms of regularization. It is able to learn complex data patterns, however, it is not robust to outliers, i.e., values that deviates significantly from the rest of data.

- L1 Regularization: The L1 regularization penalizes the absolute value of the weights according to the following:

$$L(w_{ij}) \equiv L(w_{ij}) + \lambda \sum_{i,j} |w_{ij}| = \sum_{i,p} (b_{pi} - y_{pi})^2 + \lambda \sum_{i,j} |w_{ij}| \tag{5.11}$$

  In this case, the term added to every weight is $\lambda|w_{ij}|$. It is also possible to combine this with the term from the L2 regularization, technique which is called elastic net regularization (see Zou and Hastie, 2005). The L1 regularization leads the weight vectors to be very close to zero. In other words, the network ends up using only a limited number of the most important inputs and becomes almost insensitive to noisy inputs. This behavior is opposite to that of the L2 regularization. L1 regularization is robust to outliers, however, it produces simpler and more interpretable models than L2. In practice, L2 regularization can be expected to give a better performance.

- Data augmentation: This procedure is based on the other possible approach to avoid overfitting, increasing the number of examples in the training data. It is mainly used in neural networks working with images. The size of the data is augmented using a number of techniques on the data itself. These techniques include flipping, translation, rotation, scaling, changing brightness, adding noise, etc.

- Maximum norm constraint: Large weights in a neural network are a sign of overfitting. An alternative solution to the penalization of the values of the weights proposed by L1 and L2 regularizations consists of using a weight constraint. This constraint is nothing else but an absolute upper bound on the magnitude of the weights for every neuron. The weights update is performed as usual, but if the value of a weight exceeds this chosen upper bound, it is rescaled so that its value is below the limit or between a range. Typical values of this weight constraint are on orders of 3 or 4. Unlike what happens when a penalty is added to the loss function, an explicit constraint ensures the weights of the network are small, instead of just encouraging them to be so.

- Dropout: While regularization methods like L1, L2 and max norm constraint reduce the complexity of the neural network by decreasing the values of weights, dropout modifies the structure of the network in order to reduce the number of weights. The key idea of this technique, introduced by Srivastava et al. (2014), is to randomly remove some neurons from the neural network, along with all its incoming and outgoing connections, in each iteration during training. This prevents neurons from co-adapting too much. A diagram of this technique is shown in Fig. 5.6. It can be

interpreted as sampling a neural network within the full neural network, in which the parameters of the sampled network are updated based only on the input data. Dropout is extremely effective, and can be used together with any of the previously explained methods.

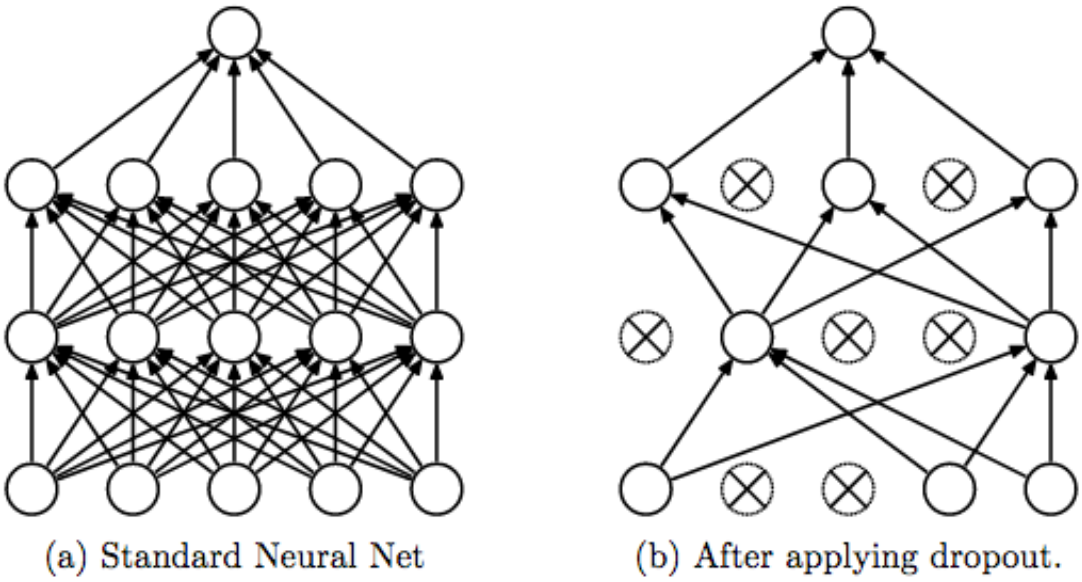


Figure 5.6: Diagram of the dropout regularization technique.
Figure obtained from Srivastava et al. (2014).

5.2 Convolutional Neural Networks

A convolutional neural network (CNN) is a type of artifical neural network designed to be applied on images, which essentially are matrices of pixel values. CNNs are the most popular deep learning architecture. They have proven to be very effective in the last years for artificial vision tasks, such as pattern recognition and image classification (Russakovsky et al., 2015). Some examples include identifying traffic signs, handwritten number recognition (Ciresan et al., 2012), facial recognition (Parde et al., 2017) or license plate detection (Kurpiel et al., 2017). The main advantage of a CNN with respect to other pattern recognition algorithms is that it automatically detects and extracts the most representative features from the images without any human supervision.

As we saw in the previous section, neural networks receive a single vector as input and operate with it through the hidden layers. Neurons within each hidden layer do not share

any connections with each other but are fully connected to all neurons in the previous layer. The output layer, which is the last fully-connected layer, represents the probability scores of belonging to the different categories in the classification problem considered. CNNs exploit the fact of receiving images as an input and, in contradistinction to regular neural networks, most of their neurons are organized in three-dimensional arrangements (width, height and depth). Neurons within these 3D layers are only connected to a small region of the previous layer, abandoning the fully-connected manner. Every layer transforms the 3D input data to a 3D output answer. In the sketch shown in Fig. 5.7, the red input layer holds the image, so its width and height have to match the dimensions of the image. On the other side, the depth will be the number of channels of the image. For instance, it would be three (red, green and blue) in the case of a color image, or higher in the case of having several astronomical images from the same extragalactic source observed at different frequencies. The final output layer will only have one dimension since the initial 3D volume of images is reduced to a single vector of class scores, settled along the depth dimension. The length of this vector will depend on the number of possible classes in the classification problem. In addition to the input and output layers, the architecture of CNNs consists of other different blocks or layers that will be detailed below.

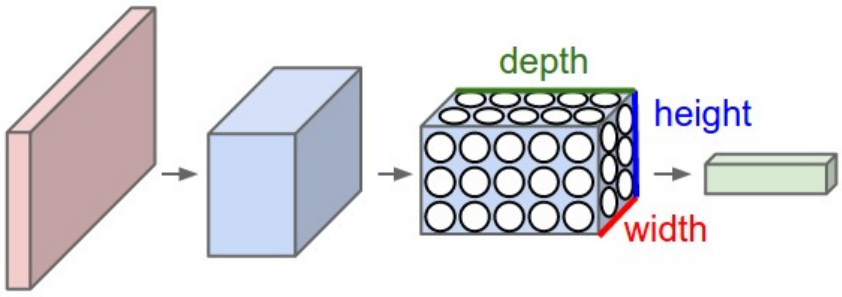


Figure 5.7: Illustration of the three-dimensional arrangement of neurons in a CNN. Taken from Fei-Fei et al. (2020).

5.2.1 Convolutional layer

The convolutional layer is the core of a CNN that does most of the computational work, as can be deduced from the fact that the entire network is named after this layer. A typical CNN usually has more than one convolutional layer. The parameters of each convolutional layer consist of a set of learnable filters, also called kernels or feature detectors, all of them being of the same size. These filters are smaller than the input images, along the height and width dimensions, but expand through the full depth of the input data. By

using small squares of the input data, the convolution preserves the spatial relationships between pixels. As can be seen in Fig. 5.8, each filter is slided, or convolved, over the width and height of the input image, computing the element-wise multiplication for every position between the entries of the filter and the pixel values of the image sub-matrix covered by the filter. The multiplication outputs for each position of the filter over the image are then added to obtain the single elements of the output matrix. This matrix, formed after sliding the filter over the image and computing the dot product, is called activation map or feature map. The process results in a two-dimensional activation map for each filter. In practice, a CNN learns the weights of these filters on its own during the training process. The network will learn filters able to recognize some kind of visual feature in the input images, such as edges, blotches or geometrical shapes. Different filters are able to detect different features from an image. The larger the number of filters, the more features can be extracted and the better the network will be at detecting complex patterns in new images.

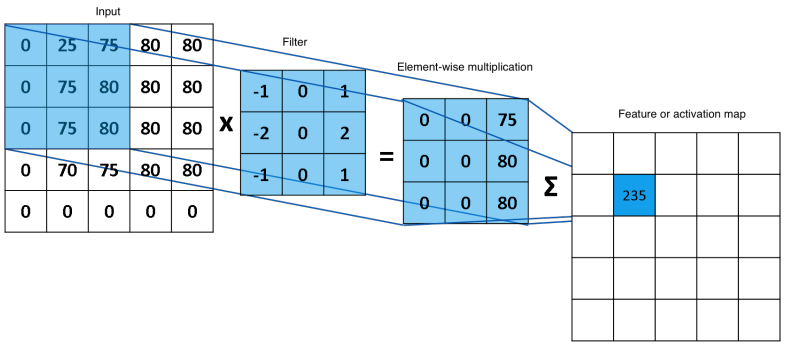


Figure 5.8: Example of the convolution operation between an input image and a filter. The filter is slided over the input image computing the element-wise matrix multiplication between it and the selected sub-matrix from the image. The values of the resulting matrix for every filter position are then added up to obtain the elements of the activation map. Taken from Robinson (2017).

The size of the activation maps is controlled by four parameters that need to be decided in advanced for every convolutional layer: the number of filters, the filters size, the stride, and the zero-padding. The number of filters per layer, also called depth, determines the number of activation maps created. If you think of them as stacked two-dimensional matrices, the number of filters will decide the depth of the whole feature map in each convolutional layer. The stride is the step used to slide the filter matrix over the input matrix. If the stride is 1, the filter scrolls one pixel at a time. If the stride is 2, the filter moves 2 pixels at a time. The larger the stride, the smaller the feature maps. The

procedure of padding the input image matrix with zeros around their borders is called zero-padding. This is convenient so we can apply the filter to the pristine input matrix, and it allows us to control the size of the activation maps.

After every convolution operation, a ReLU activation function is applied. As was explained in Sec. 5.1.3, ReLU is an element-wise operation that replaces all negative pixel values in the activation maps by zero. The purpose of using ReLU is to introduce non-linearity in the CNN since the convolution is a linear operation, but most of real data are non-linear.

5.2.2 Pooling layer

It is frequent to periodically insert a pooling layer after the convolutional layer in a CNN architecture. The function of spatial pooling, also called sub-sampling or downsampling, is to progressively reduce the dimensionality of each activation map but retaining the most important information. This allows diminishing the number of parameters and computational work in the network, thus also helps control overfitting. The pooling layer performs independently on every depth slice of the input, and downsize it spatially using a certain operation. There are different types of spatial pooling (max, average or sum pooling) depending on the type of operation used.

The input rectified activation map is split into spatial neighborhoods or windows from which the largest element (max-pooling), the average (average pooling) or the sum of all elements (sum pooling) is taken. In practice, max-pooling has been shown to work better. Since these operations are performed on small regions of the image, the pooling makes the network invariant to small transformations, translations or distorsions in the input images. This is very powerful given that the features can be detected in the images no matter where they are located. The pooling process can also be understood, analogously to convolution, as sliding with a definite stride a filter of certain size over the feature maps, where the filter will perform the chosen operation on all the pixels marked. Fig. 5.9 shows an example of max-pooling on a rectified feature map by using a 2×2 filter applied with a stride of 2. In this manner, the input map is partitioned into four regions and the maximum value is selected from each of them. As can be seen as well in Fig. 5.9, the pooling operation is applied separately to each feature map so although the size of the maps is lower, their number is the same, i.e., the depth dimension remains unchanged.

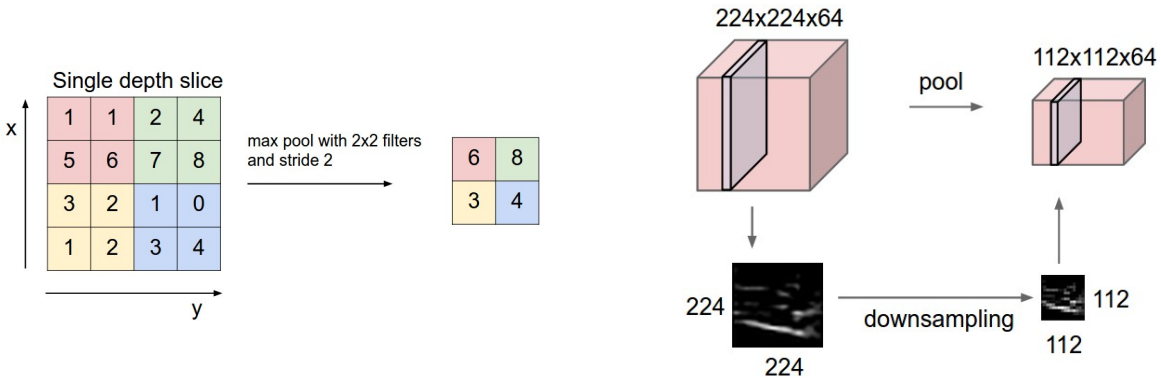


Figure 5.9: *Left*: A simple example of max-pooling where an input map of size 4×4 is split into four regions, from each of one the largest value is selected, resulting in a 2×2 output map. *Right*: An input volume of size $224 \times 224 \times 64$ is downsized with a filter of size 2×2 and stride 2 into an output volume of size $112 \times 112 \times 64$. The volume depth is conserved with only the width and height being reduced. Taken from Fei-Fei et al. (2020).

5.2.3 Fully-connected layer

So far we have explained how convolution, ReLU and pooling work. These operations together extract the helpful features from the images, introduce non-linearity in the network and reducing the dimensions of the feature maps while making the network invariant to small distorsions in the images. After the last convolution + ReLU layer, whether or not followed by pooling operation, the feature maps are flattened as a one-dimensional vector and arrive to a fully-connected layer. More than one fully-connected layer can be used in a CNN. A fully-connected layer is basically a MLP (see Sec. 5.1.1) that uses a softmax activation function. All neurons within this layer have full connections to all activation maps coming from the previous layers. These activation maps portray high-level features of the input images. The aim of the fully connected layer is to learn non-linear combinations from these features and classify the input images into the existing classes. Its output is an array of probabilities for the different possible class labels attached to the images. The class that receives the highest probability is chosen as the classification decision.

Fig. 5.10 shows an example of CNN containing all the layers explained. This network classifies every input image into four categories: cat, dog, bird or boat. When it receives a boat image as input, the network correctly gives the highest probability for the boat option (0.95) among the four possible options. The use of the softmax activation function ensures that the sum of output probabilities from the fully-connected layer is 1.

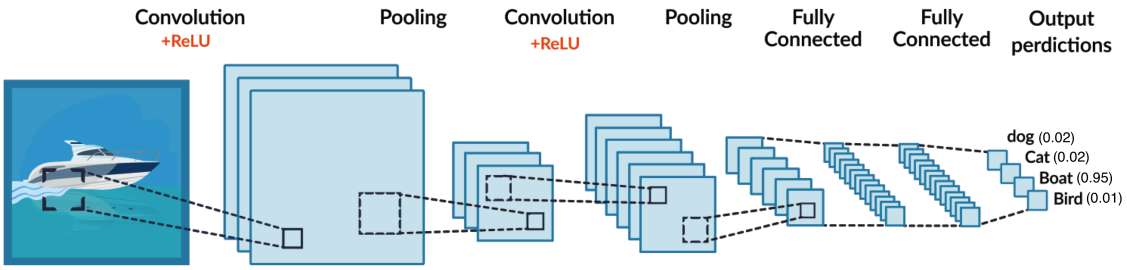


Figure 5.10: Example of CNN for classifying images among cat, dog, bird or boat. It consists of an input layer, 2 convolutional + ReLU + pooling layers, 2 fully-connected layers, and an output layer that returns the probabilities for each class. Taken from MissingLink.

5.2.4 Training process

The overall training process of a CNN follows the same steps described for an ANN in Sec. 5.1.2. In first place, the architecture of the network must be decided (number and type of layers, number of filters, filter sizes, etc). All the other parameters and weights are initialized with random values. We also have to set up how the training data will be handled by the network. This consists of deciding the batch size and the number of epochs. The batch size is the number of slots in which the training data are split while the number of epochs, or learning cycles, is the number of times that the whole training data set is passing through the CNN. The number of iterations per epoch will depend on the size of the training data set and the batch size. For example, if you have 10,000 training examples, and the batch size is 50, it will take 200 iterations to complete 1 epoch. In each epoch the order and composition of the batches change, so the network sees the examples in a different order every cycle. When this has been decided, the training images can be fed into the network, going through all its layers, until finding for each input image the probabilities of belonging to each possible class. The total error over all classes at the output layer is evaluated using the chosen loss function, for instance Eq. (5.3). The backpropagation method is used to calculate the gradients of the error with respect to all weights in the network, and the gradient descent algorithm updates all weights to minimize the error in the output layer, as shown in Eq. (5.4). Once the optimal weights are reached, we can say that the CNN has learned to classify correctly the images from the training data set and it is ready to work with a test data set.

However, sometimes the learning is not successful, and the network does not even classify the training sample well enough, or performs much better on the training data set

than in the validation data set. As it was explained in Sec. 5.1.4, this kind of learning problems can easily be diagnosed by monitoring the performance of the CNN over time during training. For classification tasks, the loss and the accuracy on both the training sample and the validation sample can be recorded for each epoch in order to inquire further into these anomalous behaviors. Any decrease in the loss, or increase in the precision, for the training data must have the same effect on the validation sample. If not, we say that the model is overfitted.

Among the techniques explained to avoid overfitting, dropout is undoubtedly the most popular regularization technique for deep neural networks. As shown in Fig. 5.6, dropout consists of assigning to all or some of the neurons in the network a certain probability p of being temporarily disabled during training time. These neurons are resampled with the same probability at every iteration, so a dropped-out neuron at one iteration can be active in the next one, and the network architecture will change at each iteration. Dropout can be used either after the convolution or pooling layers to reduce overfitting when the network has given signs of it. The other common way of reducing overfitting is to increase the number of examples in the training data set through data augmentation.

5.2.5 Performance of a CNN

The performance of any classification model, such as neural networks, on a test data set for which the true values are known can be described using a confusion matrix. The confusion matrix is a table whose number of both rows and columns matches the number of output classes from the network. It shows the amount of correct predictions for each class, as well as the amount of incorrect predictions misclassified as belonging to some of the other classes. The sum of all cells in the confusion matrix matches the size of the data set. In two-class classifiers, such as here where we try to predict whether an image is a lens or not, a threshold value is set to decide. The default value for such threshold on which confusion matrices are generally built is 0.5. Thus, all examples with probability scores greater than 0.5 are predicted to be lenses while those examples with scores lower than 0.5 are considered non-lenses. We can now define some basic terms of the confusion matrices associated to binary classifiers. All examples in the data that are true lenses will be denoted as positives (P) while all true non-lenses will do so as negatives (N). True Positives (TP) are the number of examples correctly identified as lenses. True Negatives (TN) are the number of examples correctly identified as non-lenses. False Positives (FP) are the number of examples wrongly identified as lenses. False Negatives (FN) are the number of examples wrongly identified as non-lenses. An example of the confusion matrix

that will be used in this work is shown in Table 5.1.

		Predicted	
		Non lens	Lens
True	Non lens	TN	FP
	Lens	FN	TP

Table 5.1: Example of confusion matrix.

We are now in disposition of defining some important rates which can be computed from the information of the confusion matrix. The true positive rate (TPR), also called recall or sensitivity, tells the probability of detection and is defined as the ratio between the true positives and the total positive cases:

$$Recall \equiv TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{5.12}$$

It is a proxy of the completeness of the classification. The false positive rate (FPR) informs us of the probability of false alarm and is defined as the ratio between the false positives and the total negative cases:

$$FPR = \frac{FP}{N} = \frac{FP}{TN + FP} \tag{5.13}$$

The accuracy, which indicates how many of the predictions done by the classifier are correct, is defined as:

$$Accuracy \equiv Acc = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.14}$$

And then we have the precision, which tells us the proportion of correctly identified lenses with respect to the total lens identifications done over the data. It is an indicator of the purity of the classification, given by:

$$Precision \equiv Pre = \frac{TP}{TP + FP} \tag{5.15}$$

The most common method used to quantify the execution of a classification model is the ROC (Receiver Operating Characteristics) curve (Dorfman and Alf, 1968; Powers, 2011). This curve represents the FPR versus the TPR (or $Recall$) for different probability thresholds (P_{thr}). Some examples of ROC curves can be seen in Fig. 6.4. The AUC

(Area Under the Curve) provided by this curve is the standard figure of merit used for comparying the performance of several classification models. It tells how much a model is able of distinguishing between classes. We seek to maximize the TP, and minimize the FP values. Therefore, the closer to the left x-axis and upper y-axis, the higher the AUC and the better the classifier.

Finally, once all the above quantities have been introduced we can define the F_β score:

$$F_\beta = (1+\beta^2)\frac{Pre \times Recall}{(\beta^2 Pre + Recall)} \tag{5.16}$$

So $F_{\beta=0} = Pre$, $F_{\beta=\infty} = Recall$, and $0 \leq F_\beta \leq 1$. $F_\beta = 0$ occurs if there are zero TP while $F_\beta = 1$ happens when there are no FP and no FN. The relative importance of $Precision$ and $Recall$ depends on the value chosen for β. The F_β can be calculated for a specific threshold P_{thr} on the score given by the classifier. The performance of a classifier can be then measured by the maximum F_β reached for any P_{thr}:

$$F_\beta = \max_p F_\beta(p) \tag{5.17}$$

When searching for lenses, having high precision, i.e., having low contamination rate is important due to the fact that in real data lenses are rarer than they are in simulated data by a factor of ~ 1000. We will use the F_β score to test the performance of the CNN in Sec. 6.3. We will assume the value $\beta^2 = 0.0009$, which will ensure that a high scoring classifier provides a pure sample of lenses on a test sample with real ratios of lenses to non-lenses.

5.3 Our network architecture

The CNN configuration used in this work is an adaptation of the network developed for the morphological classification of galaxies in Domínguez Sánchez et al. (2018), now aimed for the identification of strong lenses. It has been built and trained in Python 3.6 using the Keras library[1], a high-level, open-source neural networks application programming interface. The basic architecture of this CNN is shown in Fig. 5.11. It is composed of 4 convolutional layers with squared filters of different sizes (6×6, 5×5, 2×2 and 3×3, respectively), and a fully connected layer. A ReLu activation function is applied after every convolutional layer, and a 2×2 max-pooling is applied after the second and third

[1] https://keras.io/

convolutional layers. In the beginning, dropout was performed after each convolutional layer with rates of 0.25 or 0.5, as it is depicted in the scheme. However, after performing some trials with KiDS-like lensing simulations in Sec. 6.1, we decided to remove all the dropout except for that in the fully-connected layer. The network architecture remained that way from that moment.

We have used different input data for the network along this work, and accordingly the input layer has not always been the same. The input layer shown in Fig. 5.11 corresponds to the network used in Sec. 6.1, where we have trained using KiDS-like images, with 101×101 pixels, in three different bands (g, r and i). Hence, the matrices read by the network have dimensions (101,101,3). In Sec. 6.2 we trained using Euclid-like images, also with 101×101 pixels, only in the visible band, so the matrices read by the network have dimensions (101,101,1). In Sec. 6.3, we trained the network again using Euclid-like simulations, but with 66×66 pixels, and the imaging is available in four different bands (VIS, J, Y and H). The matrices read by the network this time have dimensions (66,66,4). Finally, in Sec. 7.4 the network is trained using two pseudo-spectra images, with 56×56 pixels, for every example to be classified. Therefore, the matrices to be fed into the network have dimensions (56,56,2). In all cases, the flux values of the images are normalised before entering the network so as to not working with large numbers.

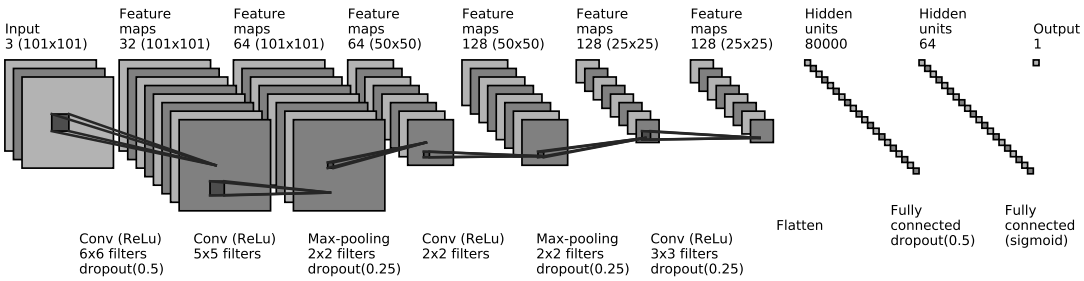


Figure 5.11: Scheme of the main architecture of the CNN used for identifying strong lenses.

The models are trained using a truth table (supervised learning) as binary classifiers, between lenses (1) and non-lenses (0). A sigmoid activation function is used in the last fully connected layer to turn the output scores into probabilities distributed between 0 and 1. In the learning process we use the binary-crossentropy loss function, defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[b_i \ln(y_i) + (1-b_i)\ln(1-y_i)] \tag{5.18}$$

where N is the number of training examples, b_i are the ground truth values, and y_i are the classification predictions of the network for each example. The stochastic gradient descent method used here is the adam optimization (Kingma and Ba, 2015) with a learning rate of 0.001. The output once the network has been trained is an array of probability scores between 0 and 1, whose length agrees with the number of examples given to the network. An score of 0 means the lowest confidence that the given example is a lens, while an score of 1 signifies the highest confidence that it is a lens. We used 25 training epochs when trying different strategies to work with the network, such as changing the size of the training samples or probing different data pre-processings. And 60 epochs, with the option for early stopping, when the ultimate training for a certain data set is carried out. The batch size is always fixed to 30. The order of the batches changes in every epoch. In the training process, several data augmentation techniques are performed, allowing the images to be zoomed in and out (0.75 to 1.3 times the original size), rotated (within 45°), flipped and shifted both vertically and horizontally (by 5%). This ensures the derived models do not suffer from overfitting since the input varies from one training epoch to the next one. The ratio of lenses/non-lenses is not necessarily balanced in the training data sets used. Taking this into account, we make sure to balance the weights assigned to the lens and non-lens classes during the training in order to avoid undesired biases in the performance of the CNN.

Chapter 6

Using convolutional neural networks to identify strong lenses

In this chapter the performance in finding strong lenses of the CNN described in Sec. 5.3 has been tested using mock data based on different astronomical surveys. More specifically, Euclid-like and KiDS-like simulations from the Euclid Strong Lensing Group have been considered. We will describe the characteristics of these simulated data, explain the process followed to train the network with them and expose the main results achieved.

6.1 Searching for strong lenses in KiDS-like simulations

6.1.1 Description of KiDS-like data

In first place, we trained the CNN using mock data mimicking the Kilo-Degree Survey[1]. KiDS (de Jong et al., 2013) is an optical multi-band imaging survey carried out from the European Southern Observatory's Very Large Telescope (ESO's VLT). In order to better understand the dark energy, the structure of galaxy halos or the evolution of galaxies and clusters, it has mapped almost 1,500 square degrees of the night sky in four broad-band filters (u, g, r, and i). These simulations formed part of the Gravitational Lens Finding Challenge 1.0, organized by the Bologna Lens Factory project[2], which took place from November 25, 2016 to February 5, 2017[3]. The simulated data were produced using

[1] http://kids.strw.leidenuniv.nl/
[2] https://bolognalensfactory.wordpress.com/
[3] http://metcalf1.difa.unibo.it/blf-portal/gg_challenge.html

GLAMER code (Metcalf and Petkova, 2014), which uses galaxies from the Millennium Simulation project[4] and real galaxies from KiDS as foreground lenses and background sources. The challenge was concentrated only on identifying galaxies that are lensed by galaxies, i.e., along with non-lensed galaxies, the simulations only contain different examples of this kind of lensing system. The simulated images are not meant to precisely mock the KiDS survey, but they were built to have the same noise levels, pixels sizes and sensitivities. All four SDSS bands (u, g, r, and i) were simulated, being used r as the reference band.

The contents of the data sets regarding the KiDS-like simulations are summarized in Table 6.1. The training set consists of 20,000 images in each band, being each image of 101×101 pixels, with a resolution of 0.2 arcseconds. Along with the images, a table is provided indicating whether or not each example is a lens. Lensing examples are labeled with '1's while non-lenses are labeled with '0's. We show some examples of lenses and non-lenses in Fig. 6.1 and Fig. 6.2, respectively. It is important to keep in mind that some objects will be classified as lenses, but the source may be too dim to be detected. That table also holds other helpful information to take into account when training the network. One information that we will use is the variable n_{pix}. This quantity measures, for each example in the training set, the number of pixels in the lensed source image that are greater than 1σ above the noise level in the r band, where σ is the standard deviation of the background noise. The test set, or challenge set, contains 100,000 candidates to be classified, i.e., 400,000 images counting all bands. These images have the same dimensions and resolution as those in the training set. 85,010 of the mock images in this set ($\sim 85\%$) were purely simulated, while the remaining 14,990 ($\sim 15\%$) used actual images taken from a preliminary sample of bright galaxies detected in the KiDS footprint. A table with additional information about the simulations is also provided for the test data set, but we do not know which examples are lenses and which are not. Among the available information we can find the aforementioned variable n_{pix} or the 'no_source' parameter, which indicates whether or not there is a source in the images. Another interesting parameter indicates which of the simualtions were built using actual KiDS images ('kids_lens_image'). Both training and challenge imaging data sets include images with artifacts and masked regions, i.e., areas with pixel values set manually to 100. After removing these cases ($\approx 13\%$), the size of the training data set falls to 17,374 examples, while the test data set retains 86,751 examples. We will refer to these data sets as *clean*. If we remove those examples with artifacts from the test data set containing only simulations with actual KiDS galaxies, we are left with a sample of 13,038 images. Further details about the challenge and how the simulations were created can be found in Metcalf et al. (2019).

[4] https://wwwmpa.mpa-garching.mpg.de/galform/virgo/millennium/

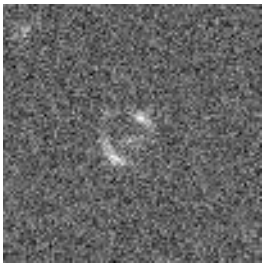
(a) ID=7 in the g-band.

(b) ID=7 in the r-band.

(c) ID=7 in the i-band.

(d) ID=268 in the g-band.

(e) ID=268 in the r-band.



(f) ID=268 in the i-band.

(g) ID=8418 in the g-band.

(h) ID=8418 in the r-band.

(i) ID=8418 in the i-band.

(j) ID=10137 in the g-band.

(k) ID=10137 in the r-band.

(l) ID=10137 in the i-band.

(m) ID=11148 in the g-band.

(n) ID=11148 in the r-band.

(o) ID=11148 in the i-band.

Figure 6.1: Some examples of lenses in the KiDS-like training data set.

(a) ID=22 in the g-band.

(b) ID=22 in the r-band.

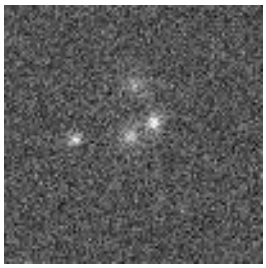
(c) ID=22 in the i-band.

(d) ID=4712 in the g-band.

(e) ID=4712 in the r-band.

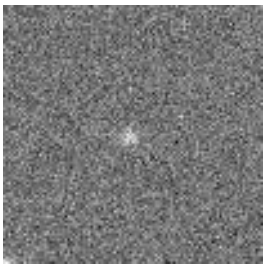
(f) ID=4712 in the i-band.

(g) ID=14338 in the g-band.



(h) ID=14338 in the r-band.

(i) ID=14338 in the i-band.



(j) ID=15515 in the g-band.

(k) ID=15515 in the r-band.

(l) ID=15515 in the i-band.

(m) ID=17776 in the g-band.

(n) ID=17776 in the r-band.

(o) ID=17776 in the i-band.

Figure 6.2: Some examples of non-lenses in the KiDS-like training data set.

Name	Description	Lenses	Non-lenses
KiDS-like training data set	20,000 images (101×101) in i, g and r bands	8,021	11,979
Clean KiDS-like training data set	17,374 images (101×101) in i, g and r bands	6,960	10,414
KiDS-like test data set	100,000 images (101×101) in i, g and r bands	49,862	50,138
Clean KiDS-like test data set	86,751 images (101×101) in i, g and r bands	43,320	43,431
KiDS-like test data set based on actual images	14,990 images (101×101) in i, g and r bands	7,500	7,490
Clean KiDS-like test data set based on actual images	13,038 images (101×101) in i, g and r bands	6,517	6,521

Table 6.1: Description of the data sets regarding the KiDS-like simulations. By *clean* we mean that those examples with artifacts and masked regions were removed from the data sets. And if the data sets just contain mock images done using actual images taken from a preliminary sample of bright galaxies detected in the KiDS footprint, we say they are *based on actual images*.

We aim to find the model that is best able to distinguish between lenses and non-lenses in the KiDS-like simulations. In the following pages, we will show the results obtained after training our convolutional neural network under different conditions in order to achieve this purpose. Except when noted otherwise, the architecture of the network used is the same as shown in Fig. 5.11. As it was our first contact with CNNs, we decided arbitrarily to train the network using only images from the g, r, and i bands. The number of trainable parameters, or weights, of the network per each layer for these simulated data are shown in Table 6.2. The output from the trained network for each example to be classified is a score between 0 and 1, 0 signifying the lowest confidence that the example is a lens and 1 signifying the highest confidence that it is a lens.

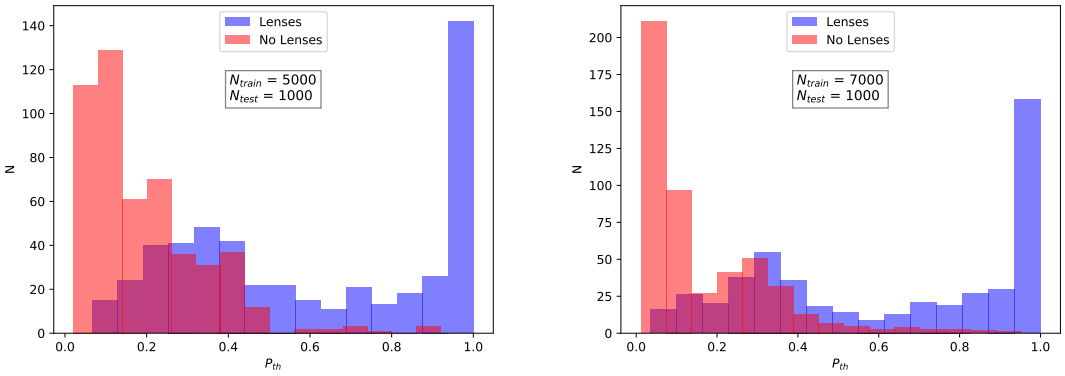
Layer	Number of weights
1st convolutional layer	3,488
2nd convolutional layer	51,264
3rd convolutional layer	32,896
4th convolutional layer	147,584
1st fully connected layer	5,120,064
2nd fully connected layer	65
Full network	5,355,361

Table 6.2: Number of weights per layer of the CNN trained with KiDS-like simulations.

6.1.2 Trials with the KiDS-like data

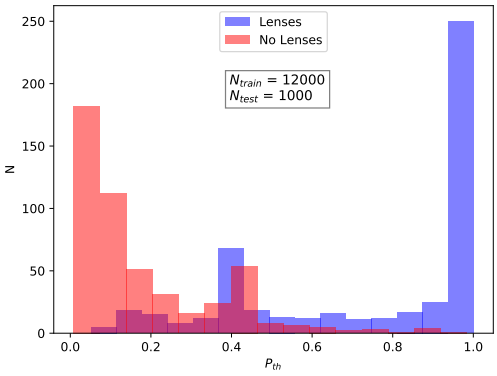
We start by running some trials in order to decide the size of the training data set (n_{train}) we will use in the different tests to be carried out. We choose four different training sizes, n_{train} = [5000, 7000, 12000, 17000], and compare the performances of the derived models with each other on the same test data set. Both the training samples and the test sample were selected from the *KiDS-like training data set* described in Table 6.1. The ratio of lenses/non-lenses in these training data sets is roughly the same as for the full training set, while the test sample has the same number of lenses and non-lenses. In each training we use validation samples whose size (n_{val}) is one tenth the corresponding training data set size, i.e., n_{val} = [500, 700, 1200, 1700]. Therefore, we are using 5500, 7700, 13200, and 18700 examples, respectively, in each training process. The architecture of the network is the same as shown in Fig. 5.11. Both here and for the rest of the tests carried out, we first plot the probability distributions obtained by applying the different models to the same test sample. Then we build the corresponding confusion matrices by choosing $P_{thr} = 0.5$ as the reference probability threshold. Thereby, only examples with $P_{thr} \geq 0.5$ would be considered as lens predictions while the rest would be identified as non-lens predictions. We also calculate the accuracy (Acc) and precision (Pre) for the models considered, taking into account this last assumption. Finally, we plot together the ROC curves of the different models being compared.

In the probability distributions of Fig. 6.3, it can be observed that the model trained with n_{train} = 7000 yields the clearest separation between the lens and non-lens classes, followed by the model trained with n_{train} = 12,000. From the confusion matrices shown in Table 6.3 we can see that the model trained with n_{train} = 5000 accomplishes the purest classification of lenses (Pre = 96.1%), i.e., few non-lenses are missclassified as lenses. However, the model trained with n_{train} = 12,000 is the one that obtains the greater number of correct lens classifications (TPR = 71%) and the highest overall accuracy (Acc = 83%) when considering also the classification of the non-lenses. Thus we could conclude that this last model is the best from the ones compared here. This is confirmed in Fig. 6.4, inasmuch such model is by far the one that achieves a higher AUC value (area = 0.936). Henceforth, unless indicated otherwise, trainings will be done with a training sample of size n_{train} = 12,000.

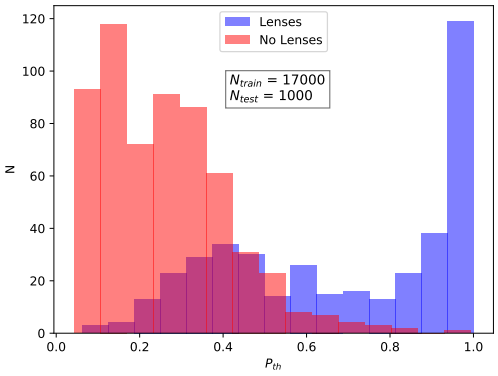


(a) $n_{train} = 5000$

(b) $n_{train} = 7000$



(c) $n_{train} = 12{,}000$



(d) $n_{train} = 17{,}000$

Figure 6.3: Probability distributions obtained by applying the models derived with the CNN using different training sizes (n_{train}) to the same test sample of size $n_{test} = 1000$, with the same number of lenses than non-lenses. Blue bins represent those examples in the test sample that are lenses while the red bins represent the non-lenses. Data were normalized in the range [-1,1].

True \ Predicted	Non-lens	Lens
Non-lens	489 (98%)	11 (2%)
Lens	231 (46%)	269 (54%)

(a) $n_{train} = 5000$
($Acc = 75.8\%$. $Pre = 96.1\%$)

True \ Predicted	Non-lens	Lens
Non-lens	476 (95%)	24 (5%)
Lens	215 (43%)	285 (57%)

(b) $n_{train} = 7000$
($Acc = 76.1\%$. $Pre = 92.2\%$)

True \ Predicted	Non-lens	Lens
Non-lens	475 (95%)	25 (5%)
Lens	145 (29%)	355 (71%)

(c) $n_{train} = 12{,}000$
($Acc = 83\%$. $Pre = 93.4\%$)

True \ Predicted	Non-lens	Lens
Non-lens	467 (93%)	33 (7%)
Lens	170 (34%)	330 (66%)

(d) $n_{train} = 17{,}000$
($Acc = 82.4\%$. $Pre = 86.8\%$)

Table 6.3: Confusion matrices obtained by applying the models derived with the CNN using different training sizes (n_{train}) to the same test sample of size $n_{test} = 1000$, with the same number of lenses than non-lenses. We chose $P_{thr} = 0.5$ as the reference probability threshold to decide whether or not an example is predicted as a lens. The accuracy (Acc from Eq. 5.14) and precision (Pre from Eq. 5.15) of the different models are also calculated based on this threshold. Data were normalized in the range [-1,1].

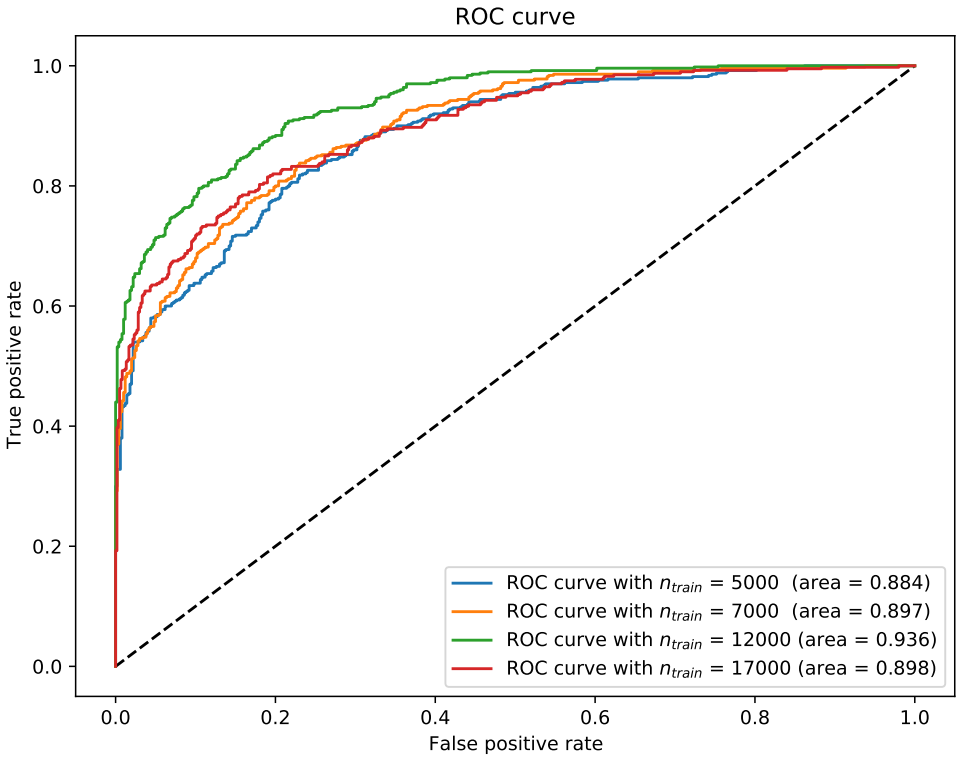
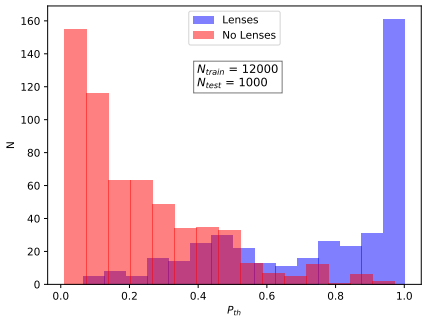


Figure 6.4: ROC curves obtained by applying the models derived with the CNN using different training sizes (n_{train}) to the same test sample of size $n_{test} = 1000$, with the same number of lenses than non-lenses. Data were normalized in the range [-1,1].

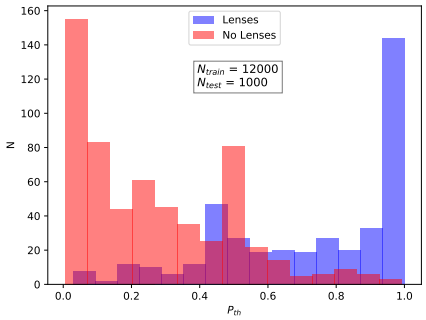
Now we proceed to evaluate the performance of the network for different normalizations of the input data. Data normalization is needed so as the network does not have to deal with very small or very large values, which could lead to problems in the learning process. In the trials shown earlier we normalized the data in the range [-1,1]. Now we also consider normalizing the data in the range [0,1], and to the maximum value in each band. This last normalization imposes an upper limit of value 1 for every example, but more freedom is granted for the lowest values, which change from one example to another, and from one band to another. Minimum values are of the order of ~ -1. We compare the performances of the models trained with the same training data, but pre-processed under these different conditions, when being applied to the same test sample. The training and test samples were both extracted from the *KiDS-like training data set*, and their sizes are $n_{train} =$ 12,000 and $n_{test} = 1000$, respectively. The training sample is the same as used before, and will be the same from now on, unless indicated otherwise. The test sample has different

number of lenses than non-lenses. The architecture of the network is the same as shown in Fig. 5.11, with 25 training epochs, and a batch size of 30.

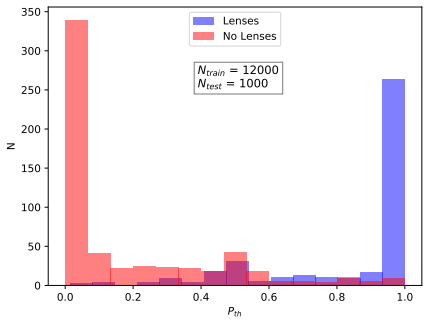
The probability distributions from Fig. 6.5 show, beyond any doubt, that normalizing the data in each example to the maximum value in each band results in the best separation between classes. The distribution of the scores for lenses clearly peaks in the valued-1 bin, while the distribution of non-lenses peaks in the valued-0 bin. The behavior between the two peaks is very flat in comparison. According to the Table 6.3, this data normalization provides the highest accuracy ($Acc = 86.5\%$), increasing significantly the number of correctly identified lenses ($TPR = 85\%$) with respect the other two models, although its precision is lower than when normalizing to [-1,1]. Nevertheless, the confusion matrix values, as well as the precision, were calculated assuming a probability threshold of $P_{thr} = 0.5$. The evident separation between classes achieved in Fig. 6.5c indicates that the precision of this model will easily overcome that other model for higher values of P_{thr}. When we plot together the ROC curves of these three models (see Fig. 6.6), we observe that the model providing the best separation between classes is also the one that achieves a higher AUC value (area = 0.953). From now on, trainings will be done normalizing the data in each example to the maximum value in each band.



(a) Normalization in the range [-1,1]



(b) Normalization in the range [0,1]



(c) Normalization to max value per band

Figure 6.5: Probability distributions obtained by applying the models derived with the CNN considering various data normalizations to the same test sample of size $n_{test} = 1000$, with different number of lenses than non-lenses. Blue bins represent those examples in the test sample that are lenses while the red bins represent the non-lenses.

True \ Predicted	Non-lens	Lens
Non-lens	538 (91%)	56 (9%)
Lens	103 (25%)	303 (75%)

(a) Normalization in the range [-1,1] ($Acc = 84.1\%$. $Pre = 84.4\%$)

True \ Predicted	Non-lens	Lens
Non-lens	516 (87%)	78 (13%)
Lens	104 (26%)	302 (74%)

(b) Normalization in the range [0,1] ($Acc = 81.8\%$. $Pre = 79.5\%$)

True \ Predicted	Non-lens	Lens
Non-lens	521 (88%)	73 (12%)
Lens	62 (15%)	344 (85%)

(c) Normalization to max value per band ($Acc = 86.5\%$. $Pre = 82.5\%$)

Table 6.4: Confusion matrices obtained by applying the models derived with the CNN considering various data normalizations to the same test sample of size $n_{test} = 1000$, with different number of lenses than non-lenses. $P_{thr} = 0.5$ is chosen as the reference probability threshold. The accuracy (Acc) and precision (Pre) are calculated.

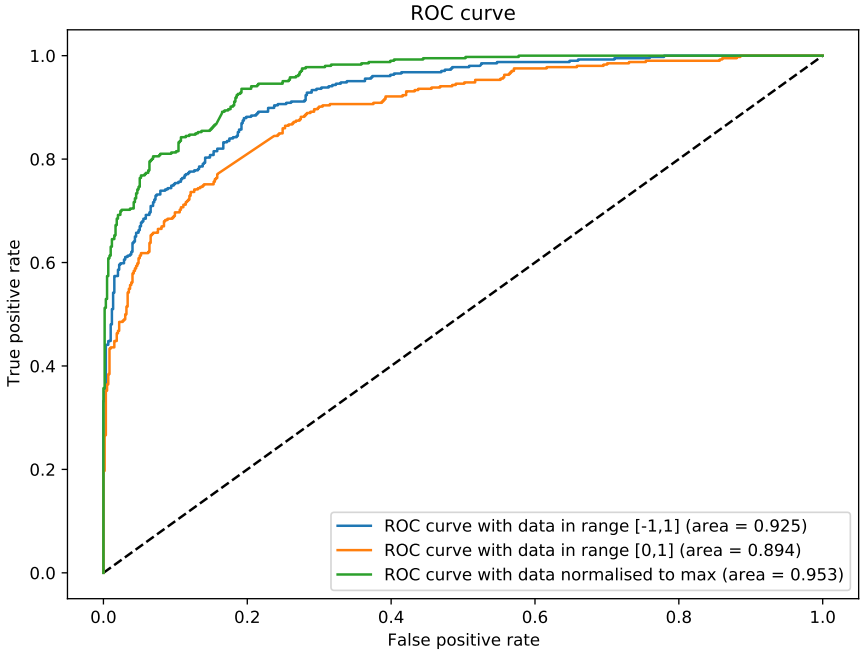
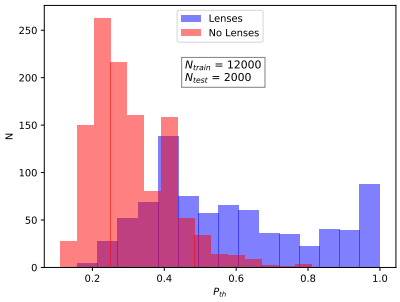


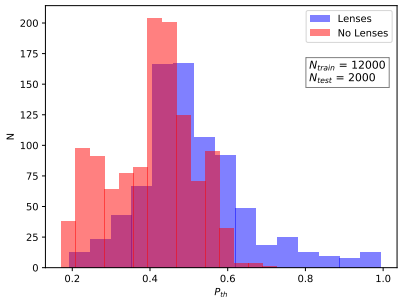
Figure 6.6: ROC curves obtained by applying the models derived with the CNN considering various data normalizations to the same test sample of size $n_{test} = 1000$, with different number of lenses than non-lenses.

In our trials working with the network we also wanted to check the impact on the learning process when modifying the input data. Here, and from now on, the test sample has been increased from 1000 to 2000 examples, extracted from the *KiDS-like training data set*. We exponentiated the data using different powers: $data^{0.8}$ (see Fig. 6.7a and Table 6.5a), $\sqrt{data}$ (see Fig. 6.7b, Table 6.5b), and $data^{1.2}$ (see Fig. 6.7c, Table 6.5c). We also forced the data to have root mean square (rms) equal to one (see Fig. 6.7e, Table 6.5e), and shift them to be all positive values (see Fig. 6.7f, Table 6.5f). The distribution of the flux values, separated by bands, of the images in the *KiDS-like training data set* is shown in Fig. 6.8. They are shown both as they are and shifted to positive values, and separated for the sake of visual clarity. A test by combining the data with a layer of gaussian noise of mean zero and standard deviation equal to 0.1 (see Fig. 6.7d, Table 6.5d) is also carried out. Adding a gaussian noise layer is useful to mitigate overfitting, and it can be seen as a form of random data augmentation. After performing every data treatment, data were normalized to the maximum value in each band. Finally, we also include an example running the CNN using raw data and normalizing them to the maximum value in each band (see Fig. 6.7g and Table 6.5g).

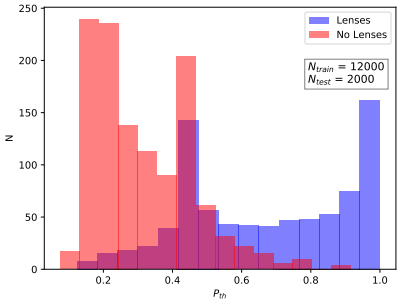
As it can be seen in the probability distributions, the model trained using data with rms = 1, and the model trained with raw data are those that achieve fairest separations between the lens and non-lens classes. The first one supplies us with the least contaminated classification of lenses ($Pre = 92.9\%$), but the latter one delivers the largest number of true lenses identified ($TPR = 79\%$) and the highest overall accuracy ($Acc = 87.3\%$). For the seak of clarity, the ROC curves from all the models compared here have been split into Fig. 6.9 and Fig. 6.10. The model trained with raw data is displayed in both figures, using a green line. This model has proven to be the best classifying the classes, since its AUC value is 0.951. It is equivalent to the model depicted, with a green line as well, in Fig. 6.6, in the sense that the training has been carried out under the same conditions. Nonetheless, both models result from different trainings, and they have been applied to different test samples. These facts sustain our conclusion that training the network without modyfing the input data, except for normalizing them to the maximum in each band, is the best choice.



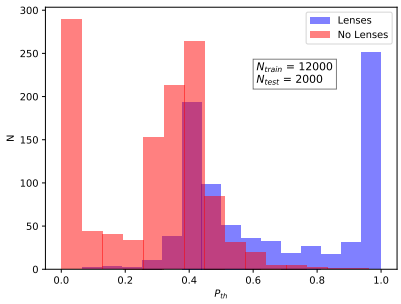
(a) $data^{0.8}$ (normalized to max)



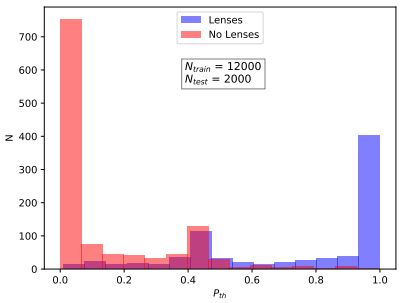
(b) $\sqrt{0.8}$ (normalized to max)



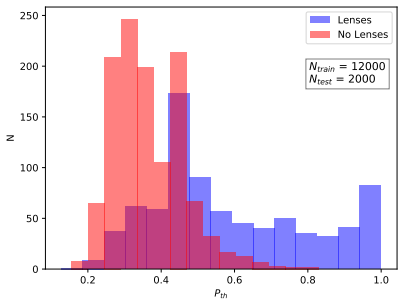
(c) $data^{1.2}$ (normalized to max)



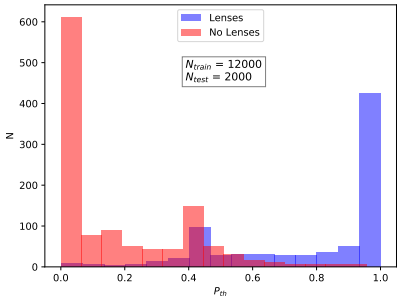
(d) data + gaussian noise (normalized to max)



(e) data with rms = 1



(f) data shifted to positive (normalized to max)



(g) raw data (normalized to max)

Figure 6.7: Probability distributions obtained by applying to the same test sample of size $n_{test} = 2000$, with different number of lenses than non-lenses, some models derived with the CNN. We considered different treatments of the data prior to being normalized to the maximum value in each band for each example. Blue bins represent those examples in the test sample that are lenses while the red bins represent the non-lenses.

True \ Predicted	Non-lens	Lens
Non-lens	1126 (95%)	61 (5%)
Lens	372 (46%)	441 (54%)

(a) $data^{0.8}$ (normalized to max)
($Acc = 78.3\%$. $Pre = 87.8\%$)

True \ Predicted	Non-lens	Lens
Non-lens	976 (82%)	211 (18%)
Lens	462 (57%)	351 (43%)

(b) $\sqrt{data}$ (normalized to max)
($Acc = 66.3\%$. $Pre = 62.5\%$)

True \ Predicted	Non-lens	Lens
Non-lens	1076 (91%)	111 (9%)
Lens	266 (33%)	547 (67%)

(c) $data^{1.2}$ (normalized to max)
($Acc = 81.1\%$. $Pre = 83.1\%$)

True \ Predicted	Non-lens	Lens
Non-lens	1114 (94%)	73 (6%)
Lens	349 (43%)	464 (57%)

(d) data + gaussian noise
(normalized to max)
($Acc = 78.9\%$. $Pre = 86.4\%$)

True \ Predicted	Non-lens	Lens
Non-lens	1144 (96%)	43 (4%)
Lens	250 (31%)	563 (69%)

(e) data with rms = 1
($Acc = 85.3\%$. $Pre = 92.9\%$)

True \ Predicted	Non-lens	Lens
Non-lens	1094 (92%)	93 (8%)
Lens	377 (46%)	436 (54%)

(f) data shifted to positive values
(normalized to max)
($Acc = 76.5\%$. $Pre = 82.4\%$)

True \ Predicted	Non-lens	Lens
Non-lens	1103 (93%)	84 (7%)
Lens	171 (21%)	642 (79%)

(g) raw data (normalized to max)
($Acc = 87.3\%$. $Pre = 88.4\%$)

Table 6.5: Confusion matrices obtained by applying to the same test sample of size $n_{test} = 2000$, with different number of lenses than non-lenses, some models derived with the CNN. We considered different treatments of the data prior to being normalized to the maximum value in each band for each example. $P_{thr} = 0.5$ is chosen as the reference probability threshold. The accuracy (Acc) and precision (Pre) of the different models are also calculated.

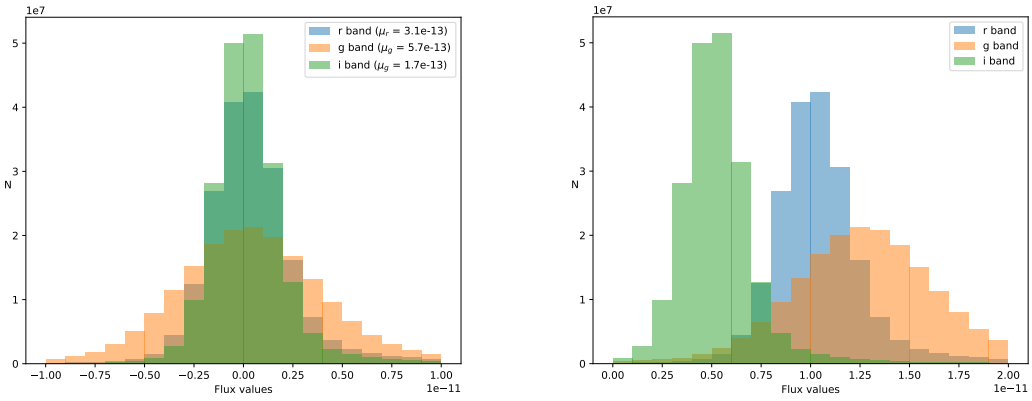


Figure 6.8: Flux distributions of the *KiDS-like training data set* for the g, i, and r bands, in which we can see the mean values have been subtracted. In the right plot, the distributions have been shifted to positive values, and separated for the sake of visual clarity.

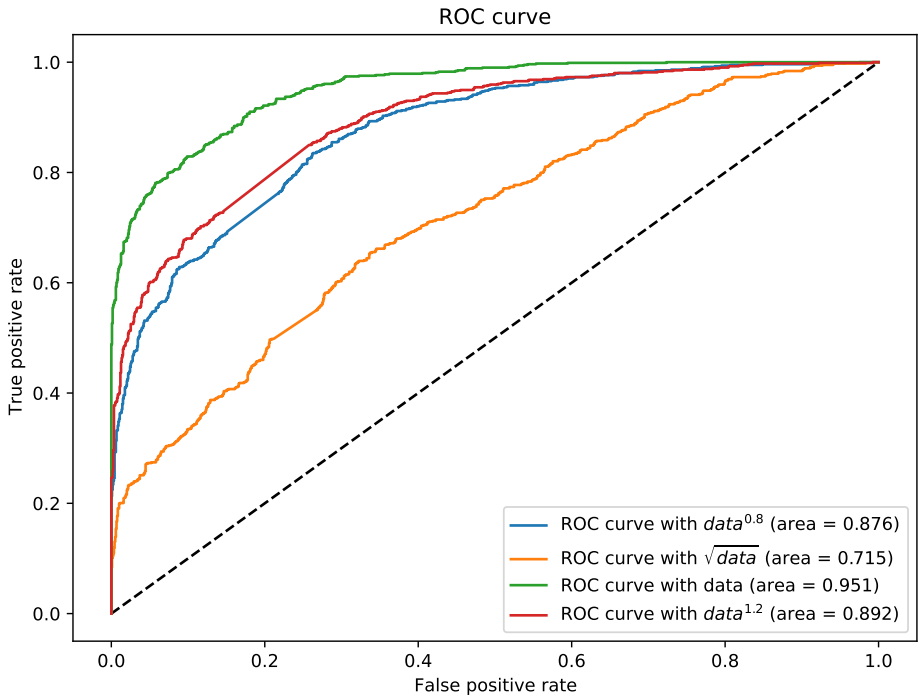


Figure 6.9: ROC curves obtained by applying to the same test sample of size $n_{test} = 2000$, with different number of lenses than non-lenses, some models derived with the CNN. We considered different treatments of the data prior to being normalized to the maximum value in each band for each example.

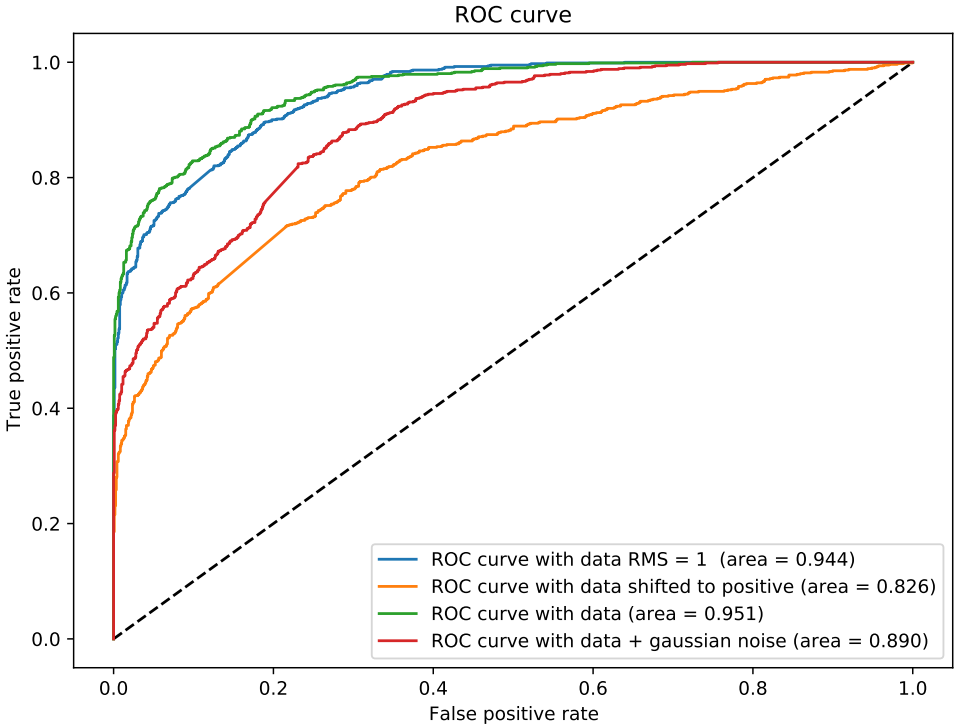
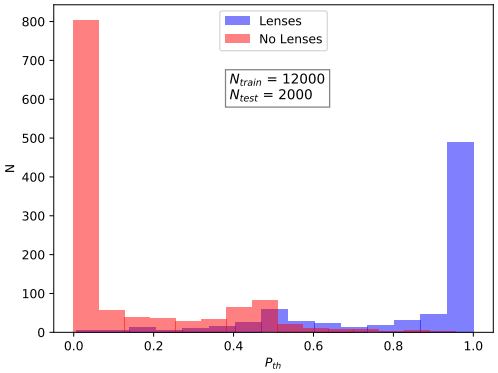


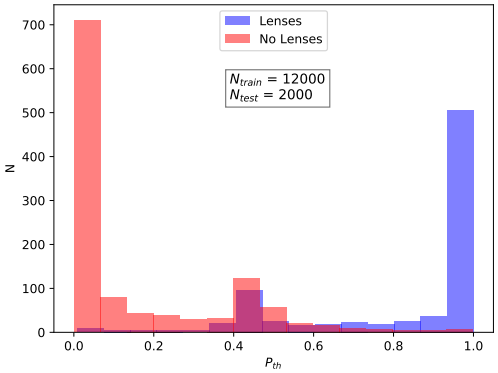
Figure 6.10: As in Fig. 6.9 but considering other data treatments. The model from Fig. 6.9 in which we train wihtout any data pre-processing (green line) is also shown here.

As we mentioned earlier, KiDS-like images do have some observational artifacts and masked regions. The extent of these artifacts varies from a few pixels to almost the entire image. Both in the training and test data sets, 4% of the examples have these masks in the three bands, 9% contain them in one or two bands, and 87% are free of them. Given that we intend to use our network with images that mimic true astronomical observations, i.e., without factitious features added afterwards, we decided to compare the performance of the network when training with and without images having artifacts. The *KiDS-like training data set* after removing such images is named *clean KiDS-like training data set*. In the same way, the *KiDS-like test data set* is reduced to the *clean KiDS-like test data set*. The characteristics of these *clean* data sets are summarized in Table 6.1. For short, we will refer to the samples containing masks as *untouched* and to the samples free of masks as *clean*. The untouched and clean training samples have the same size (n_{train} = 12,000), being the first one the same that has been used to train until now. The untouched and clean test samples also have the same size (n_{test} = 2000), being the first one the same we have used in the previous set of tests. The architecture of the network is the same as shown in Fig. 5.11, with 25 training epochs, and a batch size of 30.

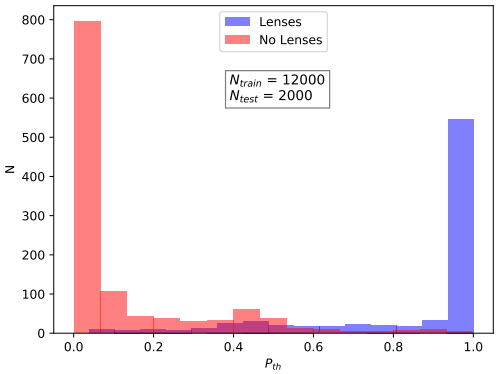
The model trained with the untouched training sample has been applied to both clean (see Fig. 6.11a and Table 6.6a) and untouched test samples (see Fig. 6.11b and Table 6.6b). The same was done regarding the model trained with the clean training sample. The probability distribution from Fig. 6.11b is the only one showing a prominent peak mixing lenses and non-lenses between 0.4-valued and 0.5-valued bins. On his side, in Fig. 6.11d it can be seen a clearly raised peak of non-lenses in the 1-valued bin, i.e., they would be misidentified as lenses with high score probabilities. We are applying here a model trained wih a clean sample to an untouched test sample. This is the only case in which the test sample contains examples the network has not seen during the training, i.e., lenses and non-lenses involving masks. Therefore, it is not surprising that the network missclassified ones or the others with high probabilities. The best separation between classes is accomplished when applying the models to a clean test sample. This can be seen both when the model has been trained with an untouched training sample (see Fig. 6.11a and Table 6.6a) as when it has been trained with a clean training sample (see Fig. 6.11c and Table 6.6c). According to Table 6.6, both tests show pretty good similar results. The quantities provided by the first one are Acc = 90.7%, Pre = 91.5%, and TPR = 85%, while the latter results in Acc = 91.1%, Pre = 91%, and TPR = 86%. The ROC curves in Fig. 6.12 show that both tests also exhibit similar behaviors, achiving a slightly superior AUC (0.974) when we have trained using a clean sample. Since only 13% of the images in the data sets carry masks, and we achieve better results training models without them, we decide that the model we will apply to the four test sets described in Table 6.1 will be trained with clean data.



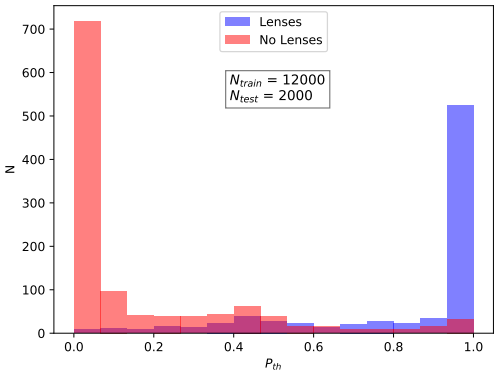
(a) Training with untouched sample, and testing with clean sample.



(b) Training with untouched sample, and testing with untouched sample.



(c) Training with clean sample, and testing with clean sample.



(d) Training with clean sample, and testing with untouched sample.

Figure 6.11: Probability distributions obtained by training and testing the models derived using different data sets. We trained the network using samples from both the *KiDS-like training data set* and *lean KiDS-like training data set*, and apply the resulting models over test samples, of size $n_{test} = 2000$, from both data sets as well. Blue bins represent those examples in the test sample that are lenses while the red bins represent the non-lenses.

True \ Predicted	Non-lens	Lens
Non-lens	1140 (95%)	63 (5%)
Lens	123 (15%)	674 (85%)

(a) Training with untouched sample, and testing with clean sample. ($Acc = 90.7\%$. $Pre = 91.5\%$)

True \ Predicted	Non-lens	Lens
Non-lens	1092 (92%)	95 (8%)
Lens	157 (19%)	656 (81%)

(b) Training with untouched sample, and testing with untouched sample. ($Acc = 87.4\%$. $Pre = 87.4\%$)

True \ Predicted	Non-lens	Lens
Non-lens	1135 (94%)	68 (6%)
Lens	110 (14%)	687 (86%)

(c) Training with clean sample, and testing with clean sample. ($Acc = 91.1\%$. $Pre = 91\%$)

True \ Predicted	Non-lens	Lens
Non-lens	1064 (90%)	123 (10%)
Lens	134 (16%)	679 (84%)

(d) Training with clean sample, and testing with untouched sample. ($Acc = 87.2\%$. $Pre = 84.7\%$)

Table 6.6: Confusion matrices obtained by training and testing the models derived using different data sets. We trained the network using data from both the *KiDS-like training data set* and *clean KiDS-like training data set*, and apply the resulting models over test samples, of size $n_{test} = 2000$, from both data sets as well. $P_{thr} = 0.5$ is chosen as the reference probability threshold. The accuracy (Acc) and precision (Pre) of the different models are also calculated.

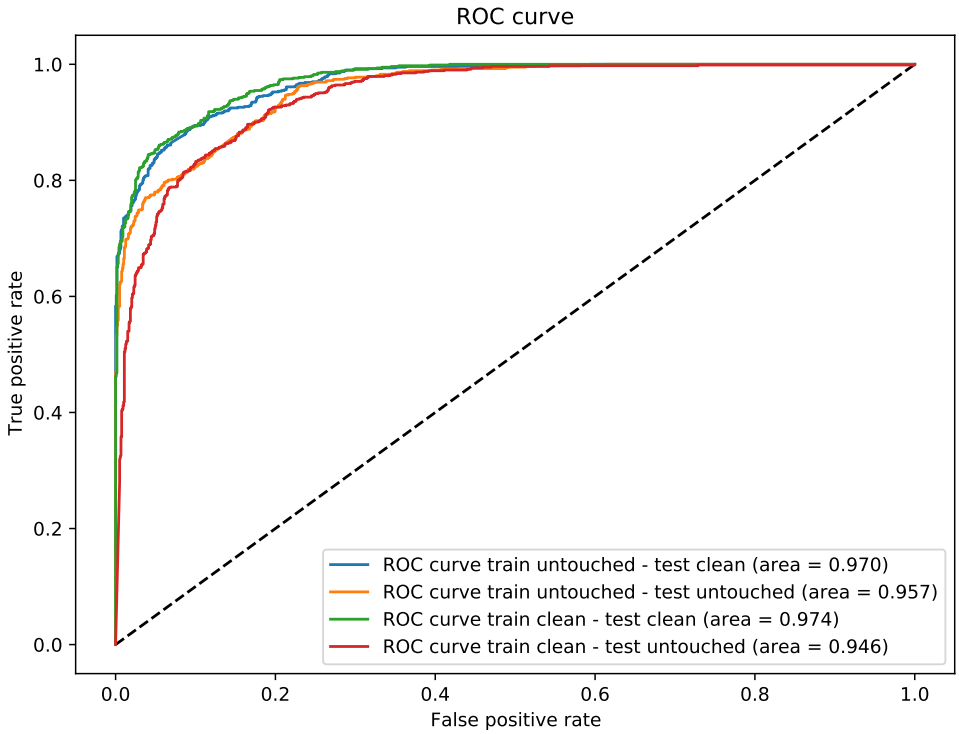
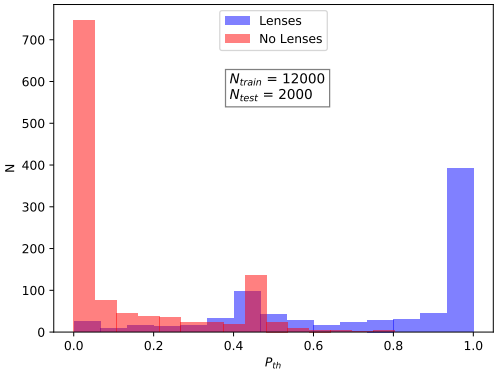
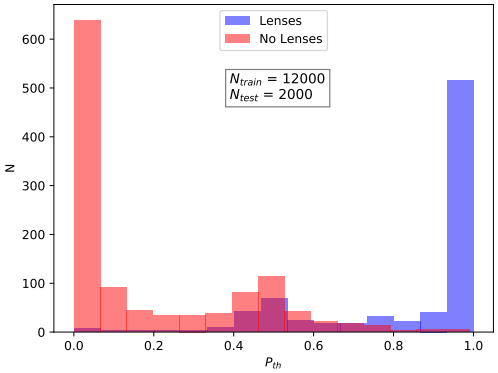


Figure 6.12: ROC curves obtained by training and testing the models derived using different data sets. We trained the network using data from both the *KiDS-like training data set* and *clean KiDS-like training data set*, and apply the resulting models over test samples, of size $n_{test} = 2000$, from both data sets as well.

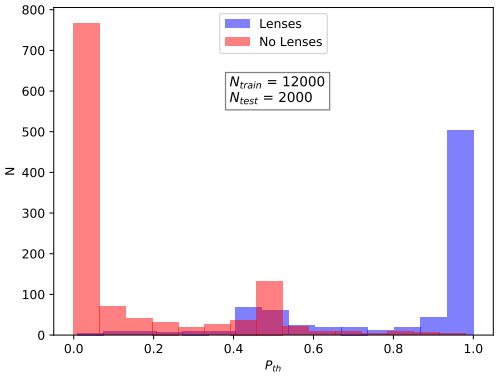
In parallel to the study of removing the examples containing masks, we have explored, as far as possible since it is beyond the scope of this work at this moment, slightly different configurations of the CNN shown in Fig. 5.11. We have tried decreasing the size of some filters or removing/adding one convolutional layer. The results were worse or in line with those obtained with the current configuration of the network. However, reducing the amount of dropouts in the network does lead to a significant improvement in the performance of the models trained. We compare in Fig. 6.13, Table 6.7, and Fig. 6.14 the results obtained both by training the network shown in Fig. 5.11, and the same network but leaving only the dropout in the fully-connected layer. These models were trained over 25 and 50 epochs, and applied to the same test sample of size $n_{test} = 2000$. As it can be seen, using less dropout, as described, results in higher overall accuracies and ROC values (0.964).



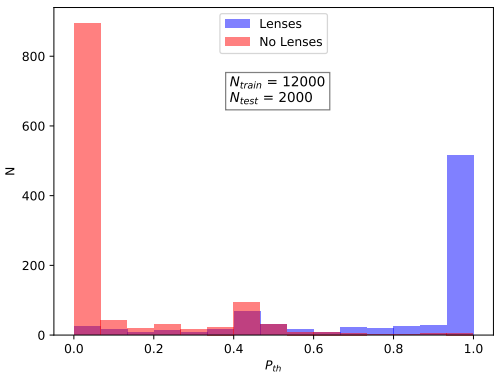
(a) All dropout (25 epochs)



(b) All dropout (50 epochs)



(c) Just last dropout (25 epochs)



(d) Just last dropout (50 epochs)

Figure 6.13: Probability distributions obtained for the models derived with two slightly different network architectures. On the one hand, dropout was performed after each convolutional layer while, on the other had, all dropout was removed except for the one in the fully-connected layer. Each case was performed running over 25 and 50 epochs. These models were applied to the same test sample of size $n_{test} = 2000$, with different number of lenses than non-lenses. Blue bins represent those examples in the test sample that are lenses while the red bins represent the non-lenses.

True \ Predicted	Non-lens	Lens
Non-lens	1150 (97%)	37 (3%)
Lens	234 (29%)	579 (71%)

(a) All dropout (25 epochs)
($Acc = 86.4\%$. $Pre = 94\%$)

True \ Predicted	Non-lens	Lens
Non-lens	1043 (88%)	144 (12%)
Lens	118 (15%)	695 (85%)

(b) All dropout (50 epochs)
($Acc = 86.9\%$. $Pre = 82.8\%$)

True \ Predicted	Non-lens	Lens
Non-lens	1111 (94%)	76 (6%)
Lens	155 (19%)	658 (81%)

(c) Just last dropout (25 epochs)
($Acc = 88.5\%$. $Pre = 89.7\%$)

True \ Predicted	Non-lens	Lens
Non-lens	1138 (96%)	49 (4%)
Lens	174 (21%)	639 (79%)

(d) Just last dropout (50 epochs)
($Acc = 88.9\%$. $Pre = 92.9\%$)

Table 6.7: Confusion matrices obtained for the models derived with two slightly different network architectures. On the one hand, dropout was performed after each convolutional layer while, on the other had, all dropout was removed except for the one in the fully-connected layer. Each case was performed running over 25 and 50 epochs. These models were applied to the same test sample of size $n_{test} = 2000$, with different number of lenses than non-lenses. $P_{thr} = 0.5$ is chosen as the reference probability threshold. The accuracy (Acc) and precision (Pre) of the different models are also calculated.

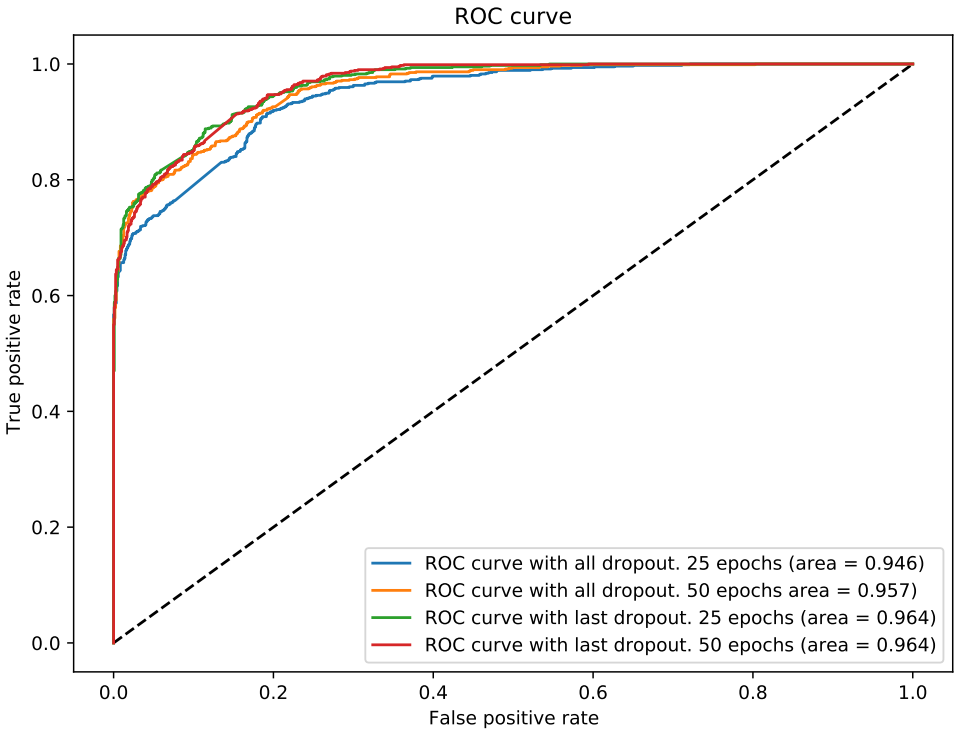


Figure 6.14: ROC curves obtained for the models derived with two slightly different network architectures. On the one hand, dropout was performed after each convolutional layer while, on the other had, all dropout was removed except for the one in the fully-connected layer. Each case was performed running over 25 and 50 epochs. These models were applied to the same test sample of size $n_{test} = 2000$, with different number of lenses than non-lenses.

6.1.3 Results for KiDS-like data

Finally, we come to the moment of applying our best model to the four test data sets described in Table 6.1. The first two data sets have already been explained. The *KiDS-like test data set based on actual images* was built by selecting in the *KiDS-like test data set* those examples that use real galaxy images taken from the KiDS footprint. We can identify them thanks to the parameter 'kids_lens_image' provided in a table associated to the simulations. The *clean KiDS-like test data set based on actual images* is built from the former after removing those examples containing artificial masks. Taking into account all the considerations seen so far regarding the performance of the CNN, our best model is achieved by training the network under the following circumstances:

- We train the network with the *clean KiDS-like training data set*. 16,374 examples are used for training while the remaining 1000 are reserved for validation.
- The input data for every example are normalized to the maximum value in each of the three bands considered (g, r, and i). No other data pre-processing was done.
- The architecture of the network is the same as shown in Fig. 5.11, but all dropout was removed except for the one in the fully-connected layer.
- We use the early-stopping option, which stops the training once the performance on the validation sample stops improving. This results in a training through 46 epochs.

Even though we do not have direct information about whether the examples in the test data sets are lenses or non-lenses, we can use some of the variables from the tables provided to infer it in a more than reasonably good way. If we look at the examples from the *KiDS-like training data set*, we see a strong correlation between the lenses and the n_{pix} variable. As it can be seen in Table 6.8, 93% of the lenses have $n_{pix} > 0$ while all non-lens examples have $n_{pix} = 0$. Therefore, this variable is a remarkable clue for labeling the lenses and non-lenses in the test data sets. When we examine the data table of the *KiDS-like test data set*, we see that in all examples with $n_{pix} > 0$ there is at least one source in the image ('no_source' = 0). After visual inspection of a random sample of examples with 'no_source' = 0 but $n_{pix} = 0$, we observe some obvious lenses with arcs and ring-shaped patterns. So, we decide to label as lenses those examples with 'no_source' = 0 (49,862) while those examples with 'no_source' = 1 are labeled as non-lenses (50,138). This is shown in Table 6.9. Our lenses/non-lenses numbers agree with those indicated in Davies et al. (2019), where they also test a CNN to identify images containing lensing systems using these same KiDS-like data sets.

Name	Lenses with $n_{pix} > 0$	Lenses with $n_{pix} = 0$	Non-lenses with $n_{pix} = 0$
KiDS-like training data set	7,491 (93%)	530 (7%)	11,979 (100%)

Table 6.8: Characteristics of the examples in the KiDS-like training data set.

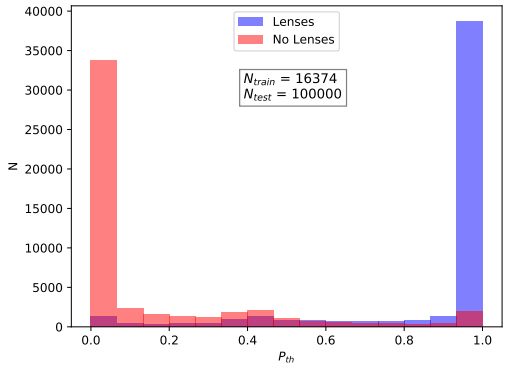
Name	no_source = 0	no_source = 0 & $n_{pix} > 0$	no_source = 1
KiDS-like test data set	49,862 (49.9 %)	46,673 (46.7 %)	50,138 (50.1 %)

Table 6.9: Characteristics of the examples in the KiDS-like test data set.

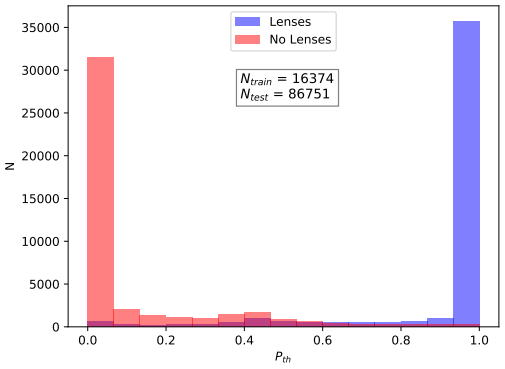
The probability distributions obtained after applying our best model to the four test data sets are shown in Fig. 6.15. We achieve a great separation between the lens and non-lens classes for the *KiDS-like test data set*, with the main inconvenient of having 2171 non-lenses missclassified as lenses with $P_{thr} > 0.9$, and 1554 lenses missclassified as non-lenses with $P_{thr} < 0.1$. These two missclassified peaks are considerably reduced, to 475 non-lenses and 729 lenses, respectively, when we consider the *clean KiDS-like test data set*. The separation between classes gets significantly worse when the model is applied just to the examples based on real KiDS images, and improves a little when the examples with artificial masks are removed. The improvement in the classification when removing the examples with artificial masks can also be appreciated in Table 6.10, Fig. 6.16, and Table 6.11. According to the values in Table 6.10, which were calculated assuming a probability threshold of $P_{thr} = 0.5$, removing these examples from the *KiDS-like test data set* translates into an increase of the TPR from 88.6% to 91.9%, the accuracy improves from 88.9% to 92.6%, and the precision moves up from 89.1% to 93.2%. In Table 6.11 one can see how this improvement remains for different probability thresholds. In this table we can also see how the stricter the probability cut, the lower is the TPR and the higher the precision, while the accuracy remains more or less stable. With high probability thresholds, we can obtain purer samples with few non-lenses missclassified as lenses (higher Pre) at the expense of reducing the number of lenses detected over all present in the data set (lower TPR). The best performance of our model is achieved when applied to the *clean KiDS-like test data set*, yielding an AUC value of 0.982 (see Fig. 6.16). Using a cut in $P_{thr} > 0.9$, we can get within this data set a sample of 39,769 examples identified as lenses with a precision of 98.7%.

We have also trained a model under the same conditions as our best model, but using only the r-band images. The training of this model took place along 57 epochs. This model was then tested on the *KiDS-like test data set*. The probability distribution is shown in Fig. 6.15e, the confusion matrix in Table 6.10e, and its ROC curve in included in Fig. 6.16. Therefore, we can do a direct comparison between this single-band model and their analogous three-band model applied on the same test data set (Fig. 6.15a, Table 6.10a). We see that the AUC value, as well as all the classification rates considered (TPR, Acc and Pre), are significantly better when the amount of information given to the network, in the form of additional bands, increases.

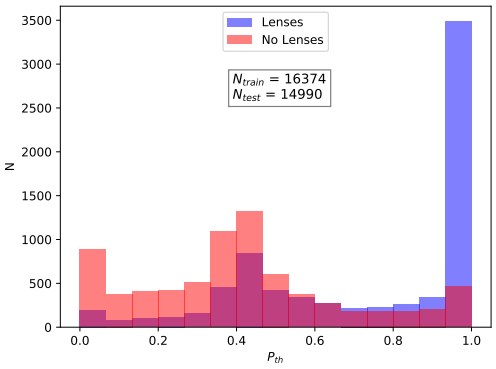
As can be clearly seen in Fig. 6.16, the efficiency of the model is very different from being applied to all the examples to being applied just to those based on actual KiDS images. The conclusion is obvious, simulations need to be as realistic as possible, so that this effect does not occur when models trained with simulations are tested on real data. Following this thought, we have also trained the network using a sample of 10,000 examples from the *clean KiDS-like test data set based on actual images* and applied the derived model to a sample of $n_{test} = 2000$ from the same data set. The resulting probability distribution and confusion matrix are included in Fig. 6.15f and Table 6.10f, respectively. The ROC curve of this model is plotted in Fig. 6.17 against the curve from our, previously explained, best model applied to the *clean KiDS-like test data set based on actual images*. The betterment now that we train and test with real KiDS images is significative. The separation of the classes is much better, being the majority of non-lenses concentrated below $P_{thr} = 0.5$. The accuracy increases in 7 percentage points whereas the precision enhances in 12 percentage points. And the AUC value moves up from 0.832 to 0.888.



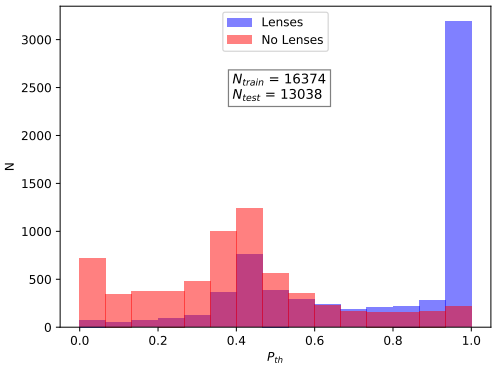
(a) *KiDS-like test data set.*



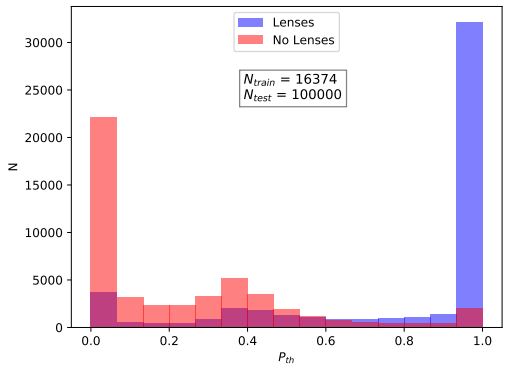
(b) Clean KiDS-like test data set.



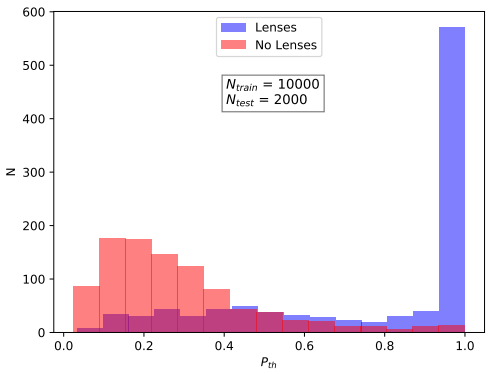
(c) *KiDS-like test data set based on actual images.*



(d) Clean KiDS-like test data set based on actual images.



(e) Training and testing with the r-band images from *clean KiDS-like test data set.*



(f) Training and testing with *clean KiDS-like test data set based on actual images.*

Figure 6.15: Probability distributions obtained by applying our best model, trained using a sample with n_{train} = 16,374 from the *clean KiDS-like training data set*, to the four test data sets described in Table 6.1 (Figures 6.15a to 6.15d). Fig. 6.15e is analogous to Fig. 6.15a, but only the r-band images were used to train the network and to test the model. Finally, Fig. 6.15f shows the probability distribution from a model trained using sample of n_{train} = 10,000 from the *clean KiDS-like test data set based on actual images* applied to a sample of n_{test} = 2000 from the same data set. Blue bins represent those examples in the test sample that are lenses while the red bins represent the non-lenses.

True \ Predicted	Non-lens	Lens
Non-lens	44725 (89%)	5413 (11%)
Lens	5692 (11%)	44170 (89%)

(a) *KiDS-like test data set* (Acc = 88.9%. Pre = 89.1%)

True \ Predicted	Non-lens	Lens
Non-lens	40542 (93%)	2889 (7%)
Lens	3509 (8%)	39811 (92%)

(b) *Clean KiDS-like test data set* (Acc = 92.6%. Pre = 93.2%)

True \ Predicted	Non-lens	Lens
Non-lens	5375 (72%)	2115 (28%)
Lens	2174 (29%)	5326 (71%)

(c) *KiDS-like test data set based on actual images* (Acc = 71.4%. Pre = 71.6%)

True \ Predicted	Non-lens	Lens
Non-lens	4859 (75%)	1662 (25%)
Lens	1736 (27%)	4781 (73%)

(d) *Clean KiDS-like test data set based on actual images* (Acc = 73.9%. Pre = 74.2%)

True \ Predicted	Non-lens	Lens
Non-lens	43279 (86%)	6859 (14%)
Lens	10700 (21%)	39162 (79%)

(e) Training and testing with the r-band images from *clean KiDS-like test data set* (Acc = 82.4%. Pre = 85.1%)

True \ Predicted	Non-lens	Lens
Non-lens	849 (87%)	127 (13%)
Lens	248 (24%)	776 (76%)

(f) Training and testing with *clean KiDS-like test data set based on actual images* (Acc = 81.3%. Pre = 85.9%)

Table 6.10: Confusion matrices obtained by applying our best model, trained with a sample of n_{train} = 16,374 from the *clean KiDS-like training data set*, to the four test data sets described in Table 6.1 (Tables 6.10a to 6.10d) Table 6.10e is analogous to Table 6.10a, but only the r-band images were used to train the network and to test the model. Finally, Table 6.10f shows confusion matrix from a model trained with a sample of n_{train} = 10,000 from the *clean KiDS-like test data set based on actual images* applied to a sample of n_{test} = 2000 from the same data set. $P_{thr} = 0.5$ is chosen as the reference probability threshold. The accuracy (Acc) and precision (Pre) of the different models are also calculated.

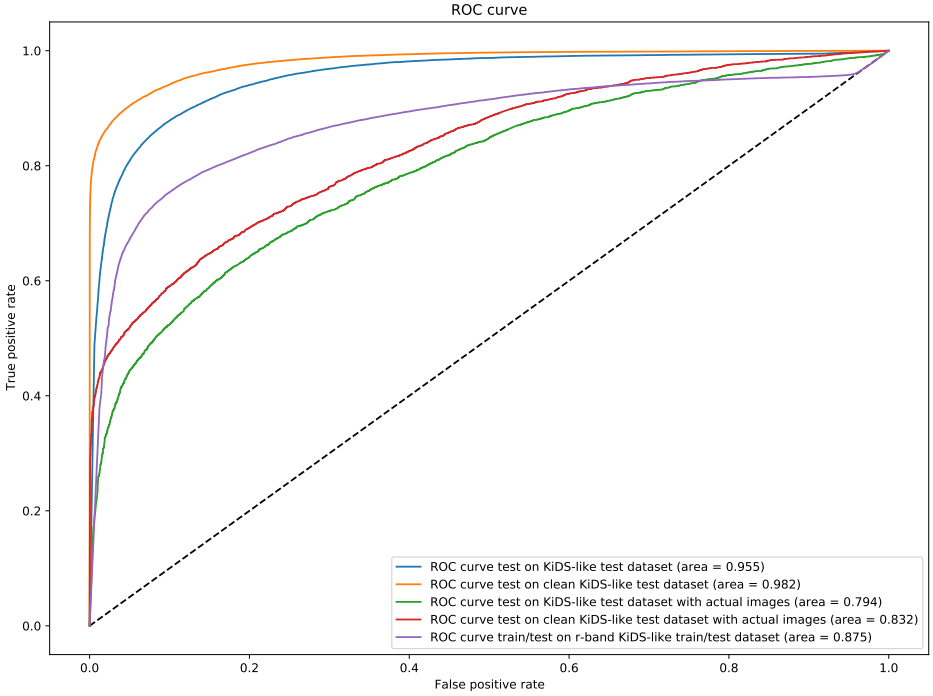


Figure 6.16: ROC curves obtained by applying our best model, trained with a sample of n_{train} = 16,374 from the *clean KiDS-like training data set*, to the *KiDS-like test data set* (blue line), *clean KiDS-like test data set* (orange line), *KiDS-like test data set based on actual images* (green line) and *clean KiDS-like test data set based on actual images* (red line). The purple ROC curve is analogous to the blue one both in training and testing, but only the r-band images were used.

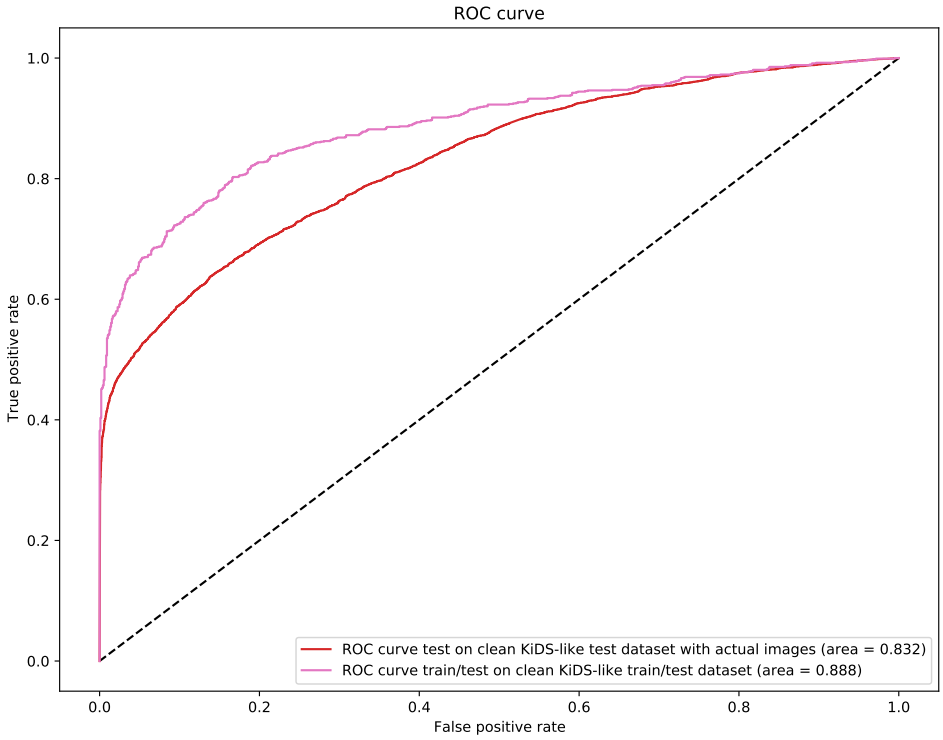


Figure 6.17: ROC curves obtained by applying our best model, trained with a sample of $n_{train} = 16{,}374$ from the *clean KiDS-like training data set*, to the *clean KiDS-like test data set based on actual images*, and a model trained with a sample of $n_{train} = 10{,}000$ from the *clean KiDS-like test data set based on actual images* to a sample of $n_{test} = 2000$ from the same data set.

Test data set	P_{thr}	TPR	Acc	Pre
KiDS-like test data set	0.5	88.6 %	88.9%	89.1%
	0.7	84.1%	88.7%	92.5%
	0.9	79.2%	87.5%	94.8%
Clean KiDS-like test data set	0.5	91.9%	92.6%	93.2%
	0.7	88.1%	92.5%	96.6%
	0.9	83.8%	91.4%	98.7%
KiDS-like test data set based on actual images	0.5	71.0%	71.4%	71.6%
	0.7	59%	72.0%	79.7%
	0.9	49.1%	70.7%	86.5%
Clean KiDS-like test data set based on actual images	0.5	73.4%	73.9%	74.2%
	0.7	61.3%	74.8%	83.8%
	0.9	51.4%	73.4%	91.6%

Table 6.11: Recall (TPR), accuracy (Acc) and precision (Pre) values for our best model trained with KiDS-like data, when applied on the test data sets and considering three P_{thr}.

6.1.4 Testing the model with strong lens candidates

We have tested the performance of our best model, trained with the *clean KiDS-like training data set*, on 101×101 pixel sized images in the g, r, and i bands of strong gravitational lenses observed with SDSS and DES. First, we have considered a sample of 174 SDSS strong lenses which has been erected as follows: 66 lenses from SLACS (Bolton et al., 2008; Shu et al., 2017), 8 secure strong lensing spiral galaxies discovered with the SWELLS Survey (Sloan WFC Edge-on Late-type Lens Survey, Treu et al., 2011), 55 lensed quasars from the SDSS Quasar Lens Search (SQLS, Inada et al., 2012), and 45 wide separation ($\geq$ 1.5 arcsec) gravitational lens systems from the CASSOWARY spectroscopy survey (Cambridge And Sloan Survey Of Wide ARcs in the skY, Stark et al., 2013). We achieve an overall precision of 65%, with 113 lenses well predicted out of 174. Those CASSOWARY lenses also confirmed as lenses by our model are shown in Fig. 6.18, while the ones our model does not identify can be seen in Fig. 6.19. The probability scores obtained for the rest of true positives from the SDSS lenses sample can be read in Table 6.12, while the false positives can be found in Table 6.13. And regarding DES, we have gathered a sample of 578 strong lens candidates discovered using CNNs, and through visual inspection. These include 503 galaxy-galaxy lenses from Jacobs et al. (2019), and 75 secure lens cadidates (those ranked with grades A and B) from Huang et al. (2020). Our model outputs a pre-

cision of 83% (481/578) for this sample of DES lenses. The probability scores obtained for all these DES strong lenses can be read in Appendix D, while some of them, both correctly and wrongly predicted by our model, are shown in Fig. 6.20. For our surprise, the model failed to identify the example J023953.1-013456.0, which is an image of the cluster A370, the first strong lensing arc discovered. We can easily find an explanation in the fact that the model we applied was trained only with examples of galaxy-galaxy strong gravitational lenses. Therefore, elongated arcs such as this one, and that in J022057.7-383303.3, are not identified by our model.

Finally, we have also applied our model on the images of two samples of potential SDSS lens candidates. On the one hand, we have drawn upon the 40 lens candidates proposed in the Table 2.4 from Sec. 2.7. 7 of them are model-predicted to be lenses. They are marked in the table with an asterisk (*), and their ID names are highlighted with orange in Appendix C. These are: J131407.0+271208 ($P_{lens} = 1.00$), J133231.6+350843 ($P_{lens} = 0.98$), J114512.9-002446 ($P_{lens} = 0.91$), J134853.0+270011 ($P_{lens} = 0.87$), J131715.6+322643 ($P_{lens} = 0.66$), J131533.9+233254 ($P_{lens} = 0.65$) and J141605.5+011828 ($P_{lens} = 0.53$). On the other hand, we have considered a sample of lens candidates from the Galaxy Zoo 2 (GZ2) project[5]. GZ2 (Willett et al., 2013) is a morphologic classification catalog of SDSS galaxies resulting from the decisions taken by volunteers through a user-friendly web interface. Even though this is an easy way to overcome the classification of large amounts of data, it can also bring about some drawbacks. The visual classification of intermediate and unclear examples can even be a difficult task for a professional. This can imply both that highly voted galaxies may not be well classified, as lead to uncertain classifications caused by the disagreement between participants. Our sample was built initially by selecting those galaxies that were largely voted for having an odd lens- or arc-type feature, according to the classification scheme followed in the project. The resulting examples were later visually inspected and labeled as lens candidates (30), unclear candidates (63), and non-lenses (116). Lens and unclear candidates were taken into account to make up a sample of 93 examples. 80 of them (28 lens candidates and 52 unclear examples) are predicted to be lenses by our model. Cutout images of all these examples are shown in Fig. 6.21.

[5] https://data.galaxyzoo.org/

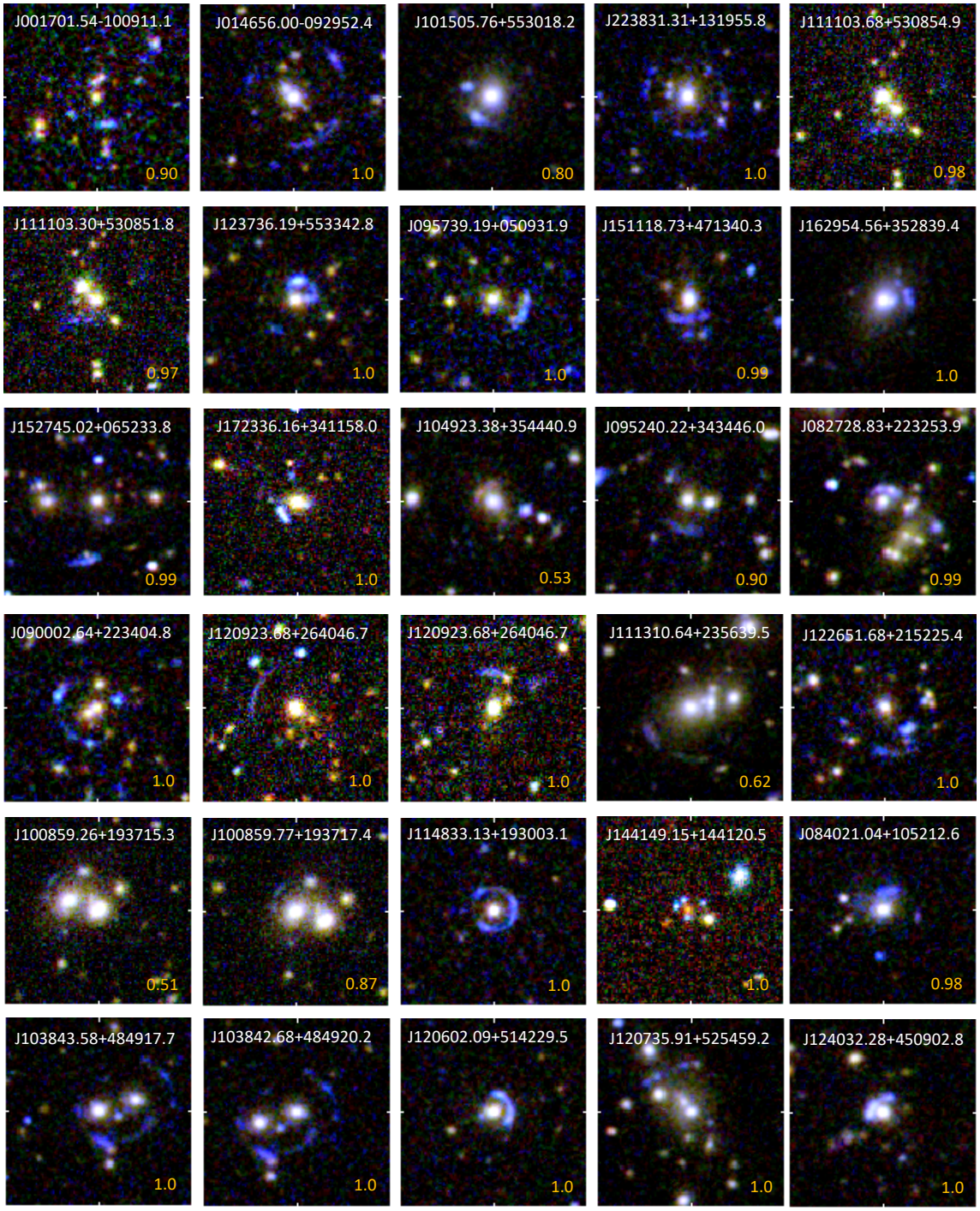


Figure 6.18: Lenses from the CASSOWARY survey (SDSS) well-predicted by a model derived with our CNN. The model was trained using the *clean KiDS-like training data set*. In orange we show the probability score of being a lens given by the model. The probability threshold to assess an example as a lens prediction is 0.5.

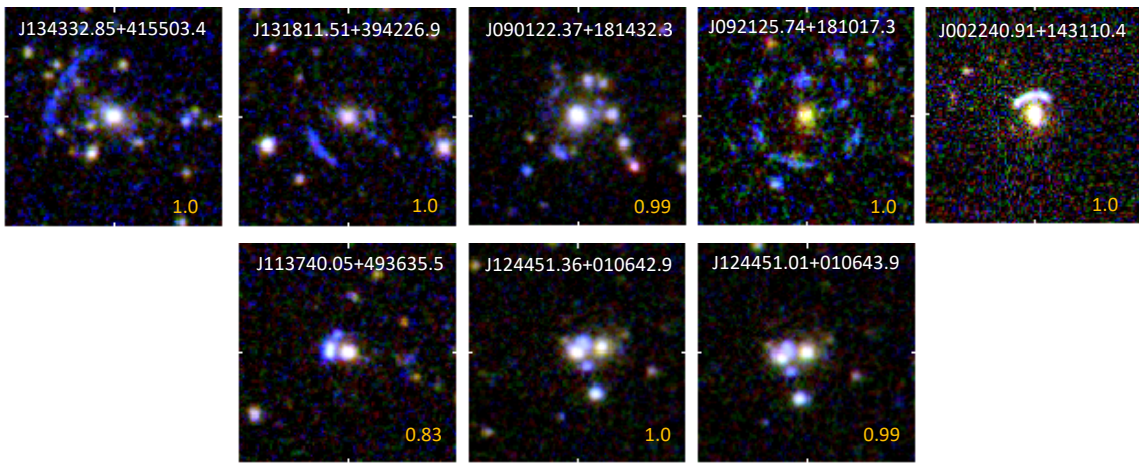


Figure 6.18 (cont.)

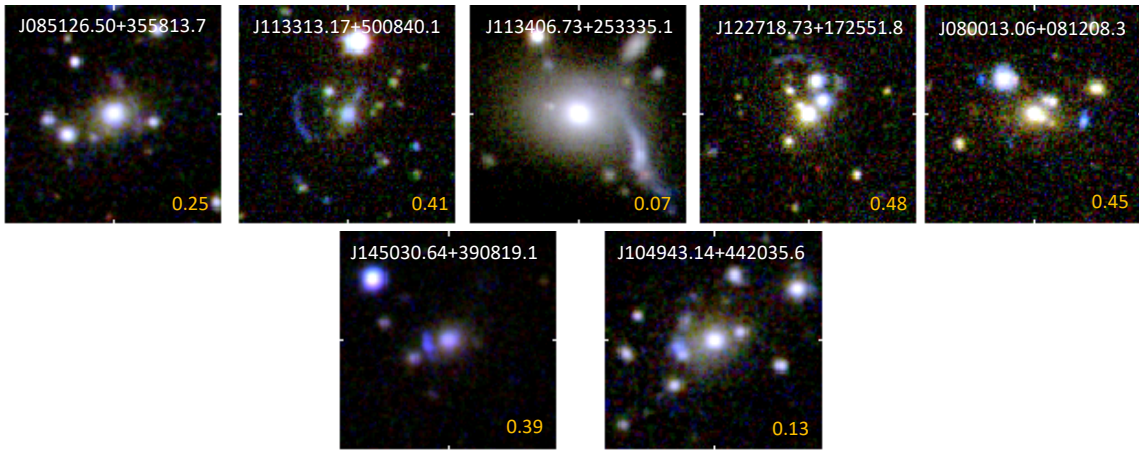


Figure 6.19: Lenses from the CASSOWARY survey (SDSS) wrongly predicted by a model derived with our CNN. The model was trained using the *clean KiDS-like training data set*. In orange we show the probability score of being a lens given by the model. The probability threshold to assess an example as a lens prediction is 0.5. After visual inspection, we do not consider the system J113406.73+253335.1 being a true lens. This is to say, the model would have been right. We are not convinced about the system J104943.14+442035.6. The rest of examples, which have higher probabilities, do seem lenses.

ID	P_{lens}	Ref
J143004.10+410557.1	1.0	[a]
J110024.39+532913.9	0.997	[a]
J095629.78+510006.6	0.987	[a]
J171837.40+642452.2	0.950	[a]
J153812.92+581709.8	0.931	[a]
J004402.90+011312.6	0.926	[a]
J140329.49+000641.4	0.924	[a]
J153150.07-010545.7	0.910	[a]
J095900.96+441639.4	0.898	[a]
J232120.93-093910.3	0.878	[a]
J125919.05+613408.6	0.870	[a]
J234111.57+000018.7	0.860	[a]
J163028.16+452036.3	0.840	[a]
J095519.72+010144.4	0.839	[a]
J125135.71-020805.2	0.809	[a]
J114257.35+100111.8	0.730	[a]
J000802.96-000408.2	0.725	[a]
J114329.64-014430.0	0.710	[a]
J003753.21-094220.1	0.673	[a]
J102922.94+042001.8	0.654	[a]
J144319.62+030408.2	0.652	[a]
J125050.52-013531.7	0.631	[a]
J101622.86+385903.3	0.545	[a]
J033012.14-002051.9	0.535	[a]
J102026.54+112241.1	0.532	[a]
J021652.54-081345.3	0.505	[a]
J142015.85+601914.8	0.501	[a]
J090334.94+502819.3	1.0	[b]
J092455.79+021924.9	0.991	[b]
J121646.04+352941.4	0.989	[b]
J100434.92+411242.7	0.989	[b]
J125819.24+165717.6	0.982	[b]
J133018.64+181032.1	0.969	[b]
J140012.77+313454.1	0.947	[b]
J125418.94+223536.5	0.941	[b]
J115517.34+634622.0	0.939	[b]
J133907.13+131039.6	0.939	[b]
J133222.62+034739.9	0.937	[b]
J091127.61+055054.0	0.927	[b]
J083216.99+040405.2	0.908	[b]
J140515.42+095931.3	0.907	[b]
J024634.09-082536.1	0.879	[b]
J122608.02-000602.2	0.871	[b]
J080623.70+200631.8	0.861	[b]
J152445.62+440949.5	0.853	[b]
J130443.58+200104.2	0.784	[b]
J135306.34+113804.7	0.778	[b]
J111816.94+074558.2	0.776	[b]
J151538.59+151135.8	0.771	[b]
J165043.44+425149.3	0.748	[b]
J120629.64+433217.5	0.735	[b]
J105545.45+462839.4	0.726	[b]
J013435.66-093102.9	0.708	[b]
J100120.83+555349.6	0.704	[b]
J115518.29+193942.2	0.688	[b]
J132236.41+105239.4	0.685	[b]
J082016.11+081215.9	0.657	[b]
J095122.57+263513.9	0.633	[b]
J105440.83+273306.4	0.606	[b]
J133534.79+011805.5	0.601	[b]
J152938.90+103803.9	0.570	[b]
J162026.14+120342.0	0.567	[b]
J125107.57+293540.5	0.552	[b]
J152720.13+014139.6	0.538	[b]
J091301.02+525928.8	0.529	[b]
J141546.24+112943.4	0.523	[b]
J113157.72+191527.7	0.506	[b]
J170341.46+245140.0	0.978	[c]
J103745.03+351730.9	0.956	[c]
J111735.37+470407.4	0.830	[c]
J082012.86+484736.7	0.813	[c]
J113506.88+372023.9	0.770	[c]
J120356.33+253549.1	0.752	[c]
J133140.33+362811.9	0.656	[c]
J091516.27+421152.7	0.636	[c]

Table 6.12: SDSS lenses well-predicted by a model derived with our CNN. The model was trained using the *clean KiDS-like training data set*. The probability score of being a lens (P_{lens}) is shown, along with the reference from which the candidate has been taken: [a] Bolton et al. (2008), [b] Inada et al. (2012) and [c] Treu et al. (2011). The probability threshold to assess an example as a lens prediction is 0.5.

ID	P_{lens}	Ref
J080858.78+470638.9	0.001	[a]
J140228.21+632133.5	0.010	[a]
J120444.07+035806.4	0.032	[a]
J002907.77-005550.5	0.040	[a]
J121826.70+083050.3	0.055	[a]
J040535.41-045552.4	0.066	[a]
J162746.45-005357.6	0.076	[a]
J090315.19+411609.1	0.076	[a]
J214154.68-000112.3	0.078	[a]
J010933.73+150032.5	0.082	[a]
J103235.84+532234.9	0.122	[a]
J113405.89+602713.5	0.132	[a]
J093543.93-000334.8	0.137	[a]
J162132.99+393144.6	0.148	[a]
J095944.07+041017.0	0.179	[a]
J082242.32+265243.5	0.222	[a]
J152506.70+332747.4	0.226	[a]
J143627.54-000029.2	0.236	[a]
J120540.44+491029.4	0.264	[a]
J102332.26+423001.8	0.274	[a]
J091205.31+002901.2	0.349	[a]
J141622.34+513630.4	0.364	[a]
J121340.58+670829.0	0.382	[a]
J111250.60+082610.4	0.384	[a]
J073728.45+321618.6	0.387	[a]
J025245.21+003958.4	0.389	[a]
J015758.94-005626.1	0.397	[a]
J125028.26+052349.1	0.399	[a]
J145128.19-023936.4	0.401	[a]
J223840.20-075456.0	0.404	[a]
J110308.21+532228.2	0.404	[a]
J230321.72+142217.9	0.408	[a]
J093600.77+091335.8	0.425	[a]
J163602.62+470729.6	0.456	[a]
J143213.34+631703.8	0.459	[a]
J230220.18-084049.5	0.472	[a]
J161843.10+435327.4	0.474	[a]
J094656.68+100652.8	0.493	[a]
J074653.04+440351.3	0.194	[b]
J100424.88+122922.2	0.196	[b]
J113803.73+031457.8	0.221	[b]
J140624.82+612640.9	0.278	[b]
J074352.61+245743.6	0.315	[b]
J094604.79+183539.7	0.385	[b]
J081331.28+254503.0	0.422	[b]
J145501.91+144734.8	0.451	[b]
J102913.94+262317.9	0.471	[b]
J133401.39+331534.3	0.474	[b]
J100128.61+502756.8	0.484	[b]
J163348.98+313411.9	0.484	[b]
J112818.49+240217.4	0.485	[b]
J132059.17+164402.5	0.494	[b]
J134929.84+122706.8	0.499	[b]

Table 6.13: SDSS lenses wrongly predicted by a model derived with our CNN. The model was trained using the *clean KiDS-like training data set*. The probability score of being a lens (P_{lens}) is shown, along with the reference from which the candidate has been taken: [a] Bolton et al. (2008) and [b] Inada et al. (2012). The probability threshold to assess an example as a lens prediction is 0.5.

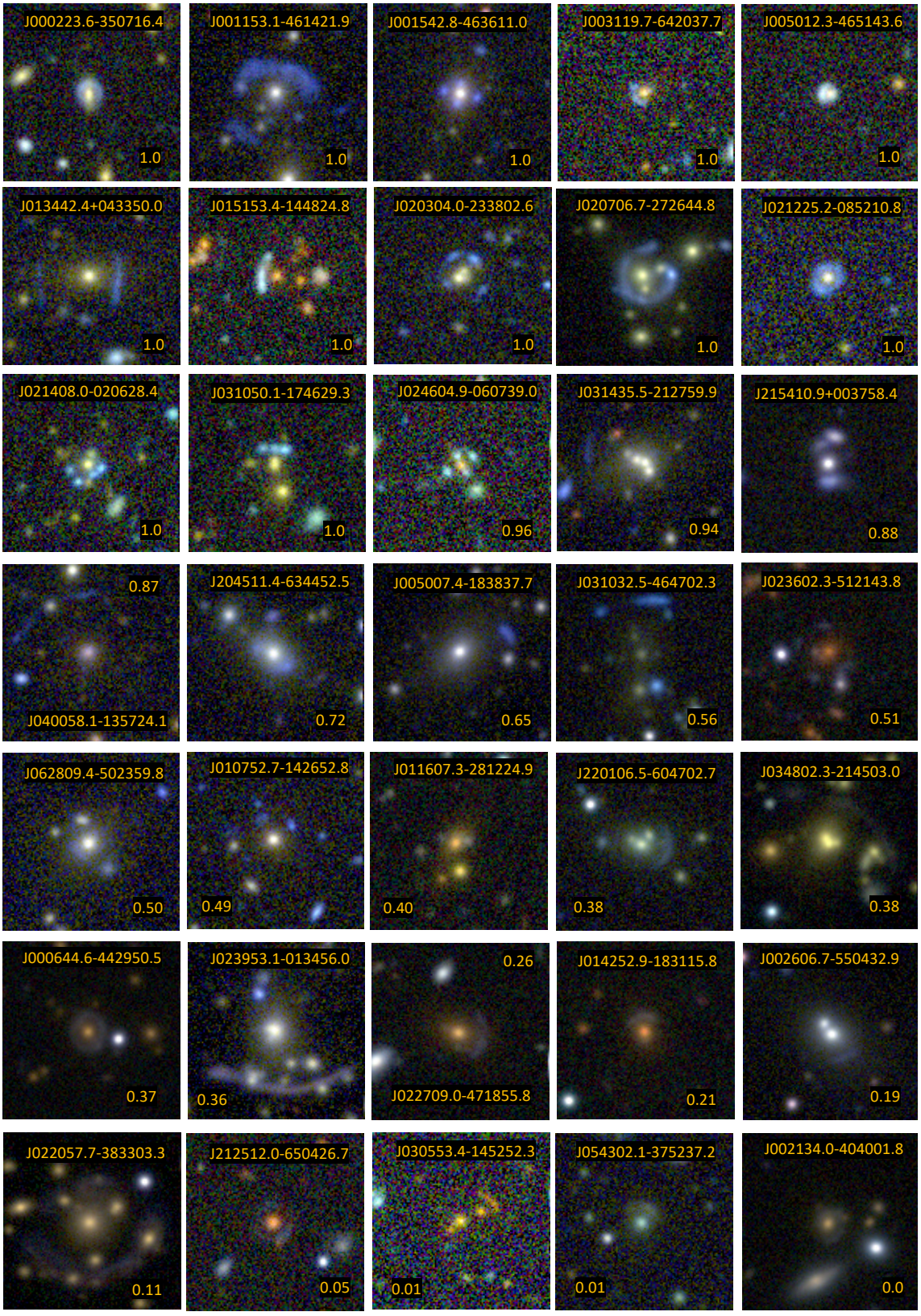


Figure 6.20: Cutouts of some DES strong lens candidates presented in Jacobs et al. (2019) and Huang et al. (2020). In orange we show the probability score of being a lens derived by the model.

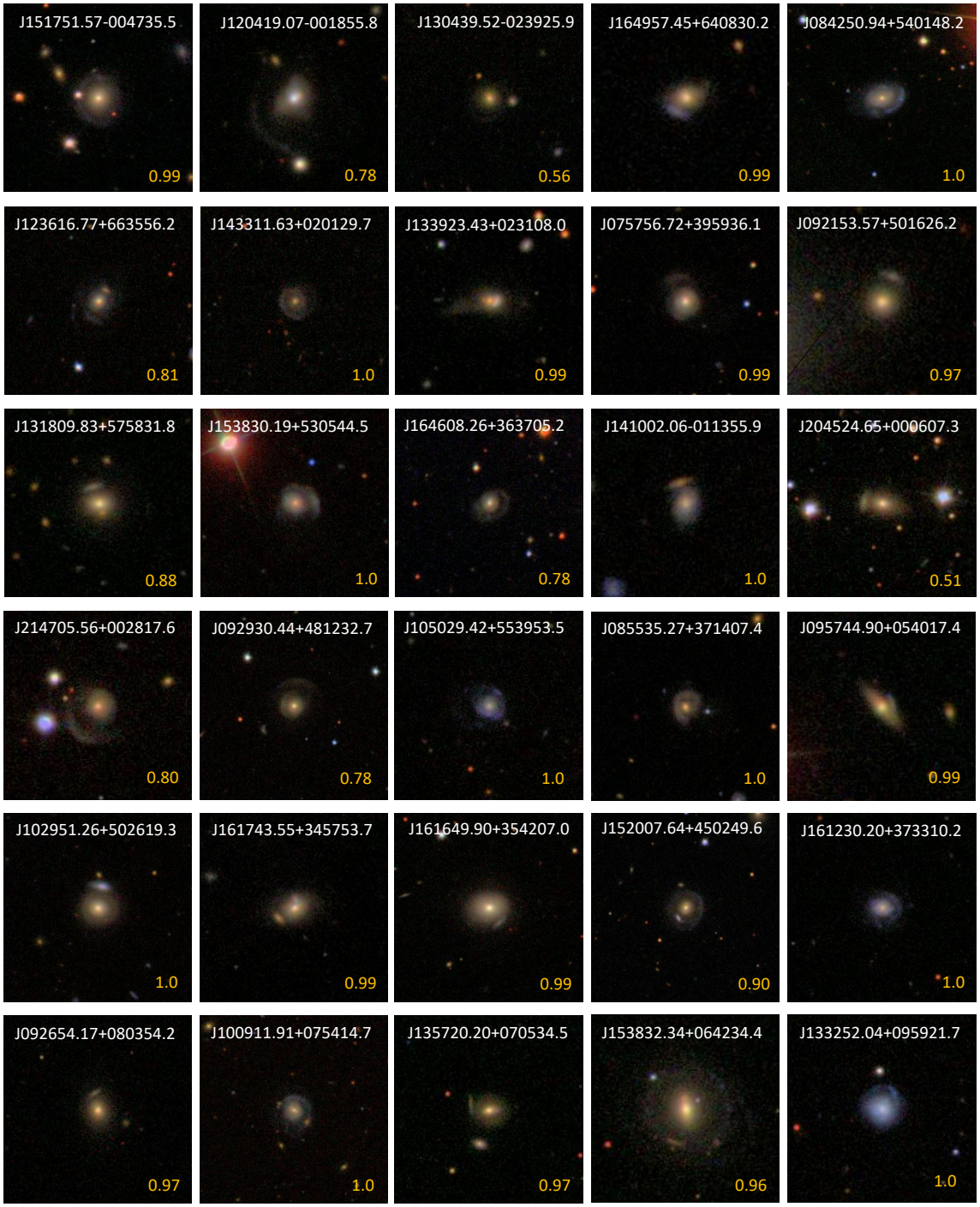


Figure 6.21: Cutouts of lens cadidates from GZ2 that are predicted to be lenses by a model derived with our CNN. The model was trained using the *clean KiDS-like training data set*. In orange we show the probability score of being a lens given by the model. The probability threshold to assess an example as a lens prediction is 0.5.

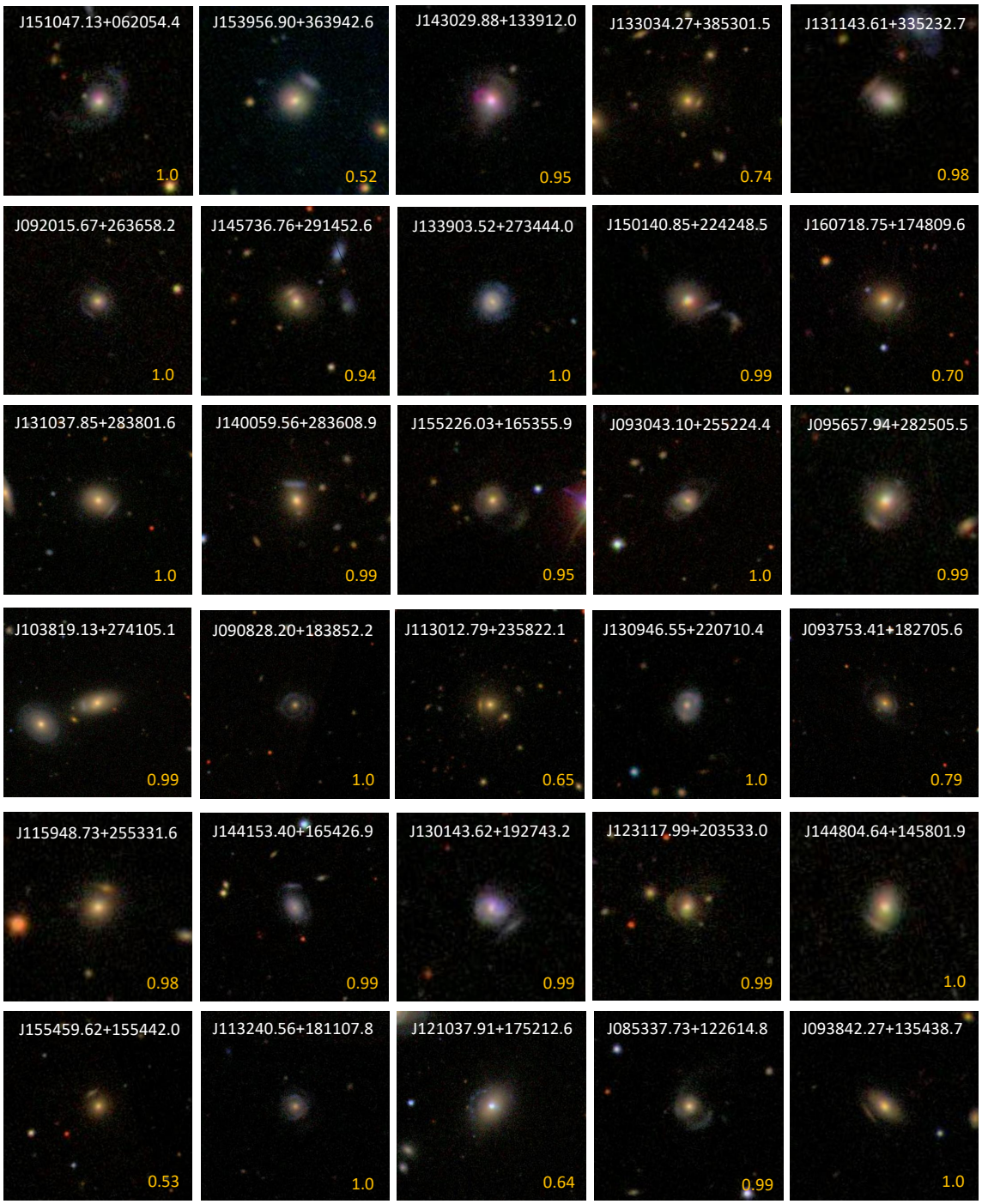


Figure 6.21 (cont.)

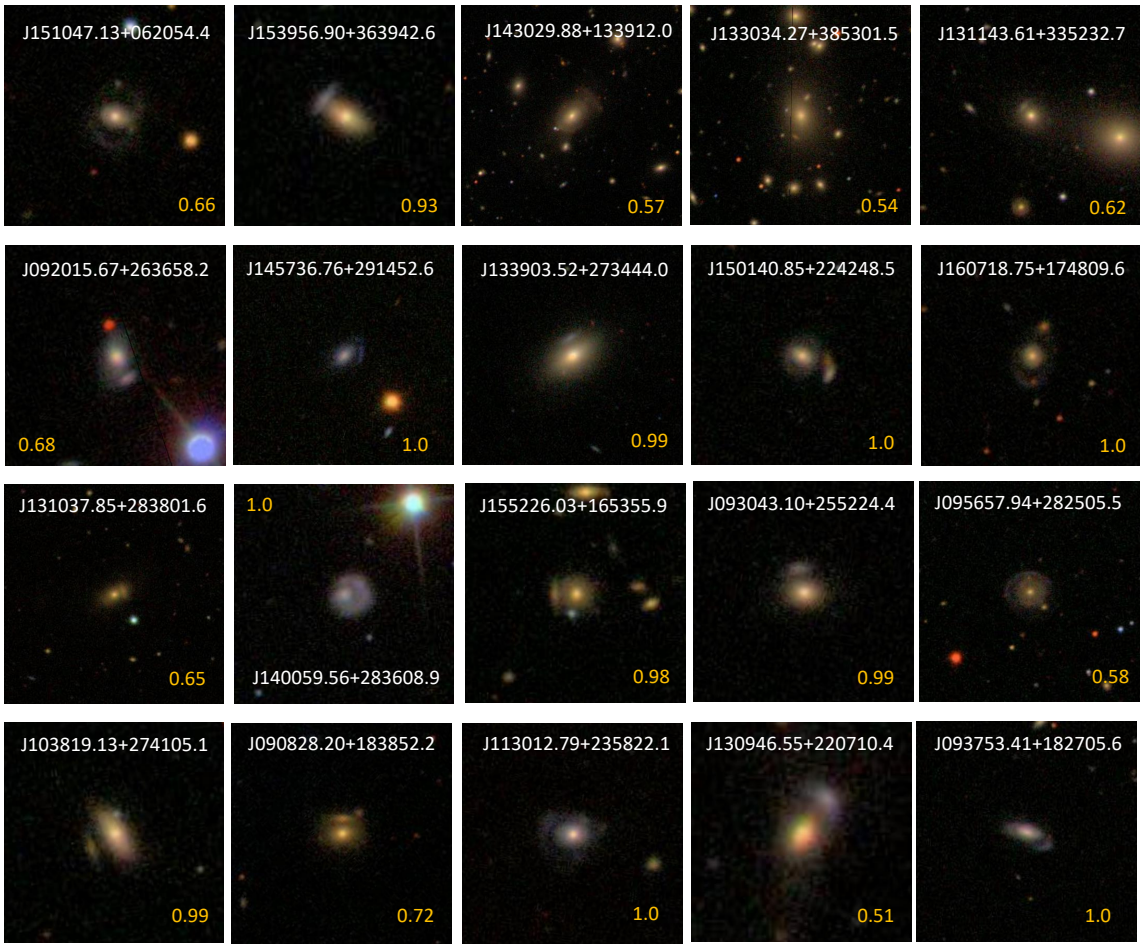


Figure 6.21 (cont.)

6.2 Searching for strong lenses in single-band Euclid-like simulations

6.2.1 Description of single-band Euclid-like data

Here we consider a data set meant to roughly mimic the data quality which is expected from observations by the Euclid telescope[6] in the visible channel. Euclid (Laureijs et al., 2011) is a space-based survey mission from the European Space Agency (ESA) designed to understand the nature of dark energy, dark matter and gravity. The Euclid wide survey will cover ∼15,000 square degrees of the extragalactic sky, complemented by two 20 deg^2 deep fields, observed in a total of four bands with two different instruments. The visible instrument (VIS) provides measurements in one broad visible $r+i+z$ band (550-920 nm), while the near infrared instrument (NISP) will image the sky in three near-infrared bands: Y (0.920-1.146 μm), J (1.146-1.372 μm) and a long H (1.372-2.000 μm). These simulations also form part of the Gravitational Lens Finding Challenge 1.0. As it occurred wih the KiDS-like simulations, only galaxy-galaxy lenses are considered (i.e., no clusters and no quasars). This imaging data set was built to have the same noise levels, pixels sizes and sensitivities as those expected from the Euclid data (see the Euclid Red Book, Laureijs et al., 2011). These space-based images have a higher signal-to-noise ratio than the KiDS-like images. The reference band for background and foreground galaxies was SDSS i, since it is overlapping with the broader Euclid VIS band.

The contents of the data sets for the Euclid-like simulations are summarized in Table 6.14. The training set consists of 20,000 images in the visible band, being each image of 101×101 pixels. Along with the images a table is provided indicating whether or not each case is a lens. Lensing examples are labeled with '1's, while non-lenses are labeled with '0's. Some examples of lenses and non-lenses are shown in Fig. 6.22 and Fig. 6.23, respectively. This table also holds other helpful information to take into account when training the network, such as the n_{pix} variable. The test set, or challenge set, contains 100,000 images to be classified. A table with additional information about the simulations is also provided. For this set we do not know which examples are lenses and which are not. The table does contain the aforementioned n_{pix} and 'no_source' variables. Further details about the challenge and how the simulations were created can be found in Metcalf et al. (2019).

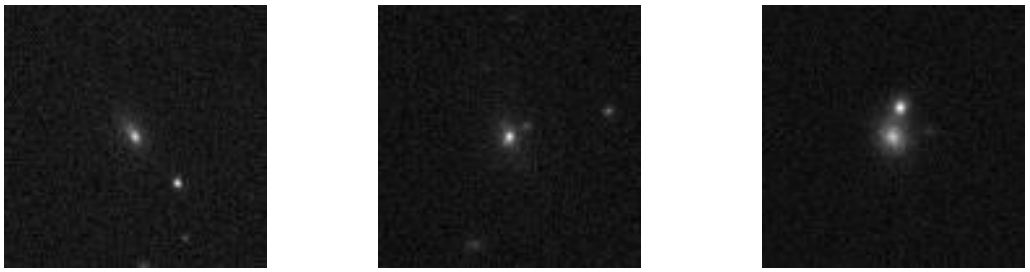
[6] https://www.euclid-ec.org/

Name	Description	Lenses	Non-lenses	$n_{pix} > 0$	$n_{pix} = 0$
Training data set	20,000 images (101×101)	13,968	6,032	13,611	6,389
Test data set	100,000 images (101×101)	39,975	60,025	38,664	61,336

Table 6.14: Description of the data sets for the single-band Euclid-like simulations.

(a) ID=19 (b) ID=65 (c) ID=6936

Figure 6.22: Some examples of lenses in the Euclid-like training data set.



(a) ID=4 (b) ID=9 (c) ID=28

Figure 6.23: Some examples of non-lenses in the Euclid-like training data set.

6.2.2 Results for single-band Euclid-like data

Bearing in mind what we learned from the trainings with the KiDS-like simulations, we proceed to tackle the lens identification within this data with three different approaches. The normalization of the input data and the architecture of the network are the same as before, with the exception that input data are now single-band. The early-stopping tool is again used here. The number of trainable parameters of the network per each layer

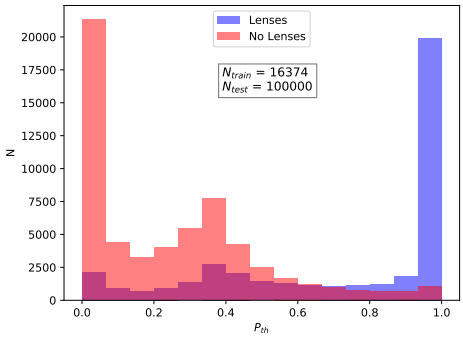
for these simulated data is shown in Table 6.15. Our first approach consisted of testing with the model from Sec. 6.1.3, which was trained using only the r-band images within the *clean KiDS-like training data set*. The training of this model lasted 57 epochs, using 16,374 examples split into 30 batches. The remaining 1000 examples were kept for validation. In the other two approaches we used the Euclid-like training data set described in Table 6.14. One model was trained making use of the ground truth labels for lenses and non-lenses provided, while in the other one we defined as lenses those examples in which $n_{pix} > 0$, and as non-lenses those with $n_{pix} = 0$. We decided to do this to check the difference between training with all the lenses and training only with the more apparent ones. 19,000 examples were used in both trainings (with 1000 for validation) and concluded after 34 and 43 epochs, respectively. These three models were all applied to the Euclid-like test data set containing 100,000 examples. As we discussed previously for the KiDS-like test data, we consider lenses those examples with 'no_source' = 0 (39,975), whereas the examples with 'no_source' = 1 are identified as non-lenses (60,025).

Layer	Number of weights
1st convolutional layer	1,184
2nd convolutional layer	51,264
3rd convolutional layer	32,896
4th convolutional layer	147,584
1st fully connected layer	5,120,064
2nd fully connected layer	65
Full network	5,353,057

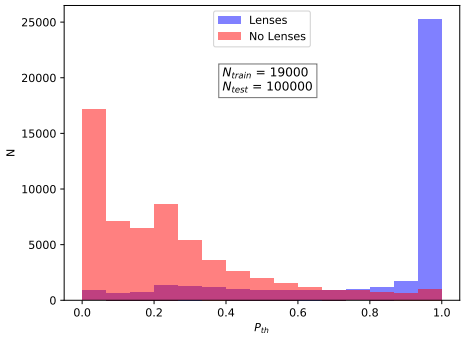
Table 6.15: Number of weights per layer of the CNN trained with single-band Euclid-like simulations.

The resulting probability distributions are shown in Fig. 6.24. Although the model trained with the r-band KiDS-like images manages to concentrate the greatest amount of non-lenses in the 1-valued bin, compared to the other models, it yields visually the worst separation between the lens and non-lens classes. Putting it into numbers, whereas this model achieves to score 52% of the lenses with $P_{thr} > 0.9$, the other two models classify 66% and 69%, respectively. On the other hand, this model misclassifies 2631 lenses (7%) as non-lenses with $P_{thr} < 0.1$, but the other two models outcome in 1254 (3%) and 568 (1.4%). This fact is also observed in the confusion matrices calculated assuming $P_{thr} = 0.5$ as border between lenses and non-lenses (Table 6.16). According to them, the model trained following the criterion $n_{pix} > 0$ to select the lenses supplies the largest number of lenses correctly identified ($TPR = 84\%$). This makes sense since we are training without

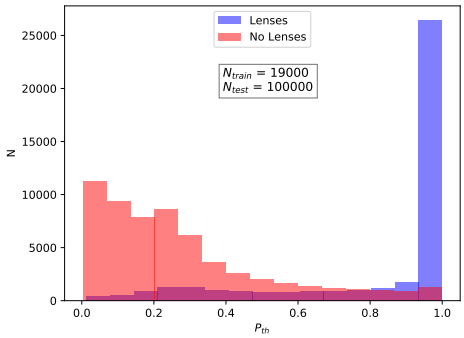
the less clear lenses, those whose number of pixels in the image are comparable to the noise level, and are few enough so we can detect best the clearest lenses, which are also a majority in the test data set. Yet the model trained with the ground truth labels delivers the least contaminated classification of lenses ($Pre = 80.4\%$). It also achieves the highest overall accuracy (84.5%), but the difference with the other model is small. As can be seen in Table 6.17, this balance is maintained with a cut in $P_{thr} > 0.9$. Training with the true lens labels we obtain the purest sample of lenses ($Pre = 95.1\%$), but training using the n_{pix} criterion renders the more accurate sample ($Acc = 85.8\%$) with the greater number of lenses correctly indetified ($TPR = 68.7\%$). The ROC curves in Fig. 6.25 also show that both models exhibit similar behaviors, but help us to confirm that being stricter in defining the lenses results in a best classifying model (AUC value = 0.921).



(a) r-band KiDS-like data (lens labels)



(b) Euclid-like data (lens labels)



(c) Euclid-like data ($n_{pix} > 0$)

Figure 6.24: Probability distributions obtained by applying three different models derived with the CNN to the single-band Euclid-like test data set of size n_{test} = 100,000. Fig. 6.24a resulted from training the network using the r-band images within the *clean KiDS-like training data set*. Fig. 6.24b shows the results of a model trained with the Euclid-like training data set using the ground truth labels given, while in Fig. 6.24c the model was trained defining as lenses those examples in which $n_{pix} > 0$. Blue bins represent those examples in the test sample that are lenses while the red bins represent the non-lenses.

True \ Predicted	Non-lens	Lens
Non-lens	51796 (86%)	8229 (14%)
Lens	11629 (29%)	28346 (71%)

(a) r-band KiDS-like data (lens labels)
($Acc = 80.1\%$. $Pre = 77.5\%$)

True \ Predicted	Non-lens	Lens
Non-lens	52162 (87%)	7863 (13%)
Lens	7627 (19%)	32348 (81%)

(b) Euclid-like data (lens labels)
($Acc = 84.5\%$. $Pre = 80.4\%$)

True \ Predicted	Non-lens	Lens
Non-lens	50580 (84%)	9445 (16%)
Lens	6509 (16%)	33466 (84%)

(c) Euclid-like data ($n_{pix} > 0$)
($Acc = 84\%$. $Pre = 77.9\%$)

Table 6.16: Confusion matrices obtained by applying three different models derived with the CNN to the single-band Euclid-like test data set of size $n_{test} = 100{,}000$. Table 6.16a resulted from training the network using the r-band images within the *clean KiDS-like training data set*. Table 6.16b shows the results of a model trained with the Euclid-like training data set using the ground truth labels given, while in Table 6.16c the model was trained defining as lenses those examples in which $n_{pix} > 0$. $P_{thr} = 0.5$ is chosen as the reference probability threshold. The accuracy (Acc) and precision (Pre) of the different models are also calculated.

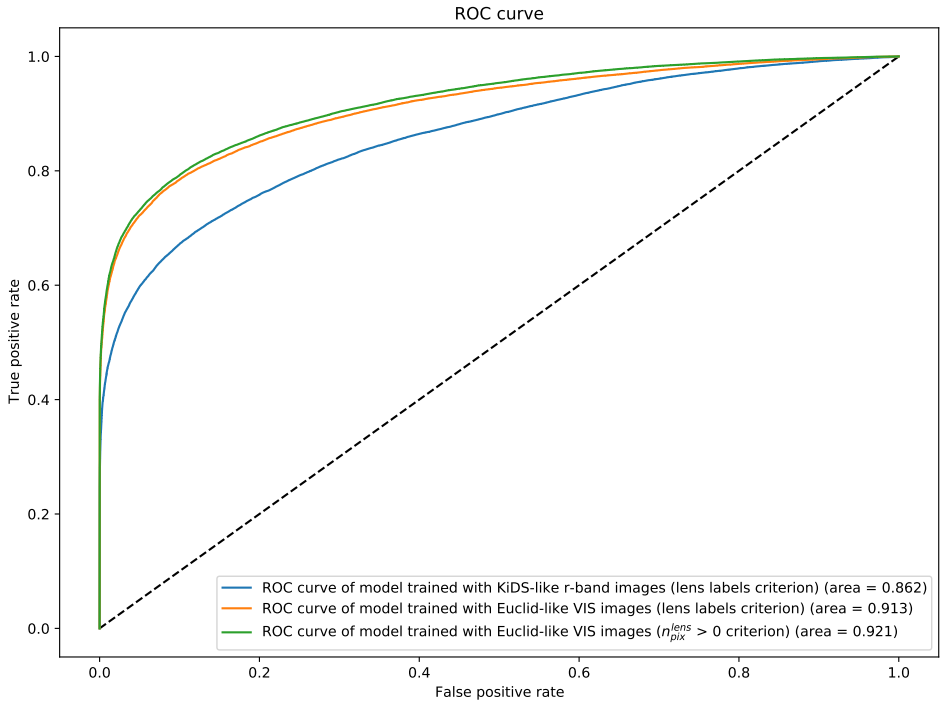


Figure 6.25: ROC curves obtained by applying three different models derived with the CNN to the single-band Euclid-like test data set of size n_{test} = 100,000. The blue line corresponds to a model trained using the r-band images within the *clean KiDS-like training data set*. The model resulted from training the network with the Euclid-like training data set using the ground truth labels given is depicted by the orange line while the green line represents a model trained with the same data but defining as lenses those examples in which $n_{pix} > 0$.

Test data set	P_{thr}	TPR	Acc	Pre
model trained with KiDS-like r-band images (lens labels)	0.5	70.9%	80.1%	77.5%
	0.7	61.9%	81.0%	86.9%
	0.9	52.3%	79.5%	93.7%
model trained with Euclid-like VIS images (lens labels)	0.5	80.9%	84.5%	80.4%
	0.7	74.0%	85.8%	88.7%
	0.9	65.6%	84.9%	95.1%
model trained with Euclid-like VIS images ($n_{pix} > 0$)	0.5	83.7%	84.0%	77.9%
	0.7	77.4%	86.1%	86.3%
	0.9	68.7%	85.8%	94.1%

Table 6.17: Recall (TPR), accuracy (Acc) and precision (Pre) values of the models applied to multi-band Euclid-like data considering three different P_{thr} cuts.

6.3 Searching for strong lenses in multi-band Euclid-like simulations

6.3.1 Description of multi-band Euclid-like data

The Bologna Lens Factory project lauched a second gravitational lens finding challenge on October 11, 2019[7], on which we entered to participate. It lasted until February 7, 2020. The simulated images within this challenge supposed an improvement over data from the previous challenges. On this occasion, lensing simulations do not limit to a galaxy being lensed by another galaxy, but multiple lensed systems involving clusters and galaxies are considered. The challenge concentrated only on Euclid-like observations in the visible, and near-infrarred J, Y, and H bands. Both the training data set and challenge data set consisted initially of 100,000 images in each of the four bands. These images are centered on a lens galaxy candidate. Visible images are 200 × 200 pixels with a resolution of 0.1". Infrarred images are 66 × 66 pixels with a resolution of 0.3". In order too work with the four images together, visible images were repixelized so they all have 66 × 66 pixels.

The training data set comes along with a catalog that gives the properties of each lens candidate. Among these properties we can find the redshifts of the source and the lens, the number of pixels in the lensed source image above 1σ (n_{pix}), the number of separated groups of source pixels, i.e., the number of images (n_{im}), the effective magnification of the source including all images (μ_{eff}), the number of sources added ($n_{sources}$) or the average surface brightness contrast between the lens and source in pixels above certain threshold ($sb_{contrast}$). The effective magnification is calculated as $\mu_{eff} = \sum_i f_i / \sum_i \mu_i^{-1} f_i$, where the sums are over the sampled pixels in the image, f_i is the surface brightness, and μ_i is the magnification of a specific pixel. A specific parameter telling us whether or not the candidates are lenses was not included. However, the organizers indicated that lenses could be defined as those examples with $n_{im} > 0$, $\mu_{eff} > 1.6$, and $n_{pix} > 20$. On the other hand, cases with $\mu_{eff} < 1.0$ or no source added are considered non-lenses. Entries for the challenge were evaluated by calculating the maximum F_β score reached for any probability threshold, according to Eq. (5.17), and assuming $\beta = 0.03$. Since a small value for β was chosen, the evaluation method rewards those models that achieve a higher precision, as defined in Eq. (5.15). Identifying all existing lenses is not as important as the fact that those predicted to be lenses are indeed lenses. Taking this into account, and due to the fact that our initial training of the network, defining the lenses and non-lenses according to the aforementioned criteria, did not succeed, we decided to use in the training a stricter

[7] http://metcalf1.difa.unibo.it/blf-portal/gg_challenge.html

definition for the lenses. Therefore, our approach consisted of training with those examples that are more likely to be clear lenses or non-lenses, leaving out those candidates which are unclear or might be hard to detect. Following this path we expected to achieve lens classifications as pure as possible in the challenge data set.

The training data set was split into two data sets, one aimed to training and the other one purposed to test our models before being applied to the challenge data set. We contemplated seven different criteria to consider the candidates as lenses and non-lenses. These criteria can be read in Table 6.18. According to them, we built six training sub-data sets and seven test sub-data sets. The criterion indicated by the organizers, nicknamed as 'assessable 0' in the tables, was not used for training, but it was used for testing. Those examples that cannot be considered lenses nor non-lenses were excluded. Regarding the challenge data set, the lenses were in the end defined using the 'assessable 1' yardstick, while all the remaining examples were labeled as non-lenses. Those examples with $1 < \mu_{eff} < 1.6$ were removed, leaving the challenge data set with 88,910 candidates. The contents of all these data sets are summarized in Table 6.19.

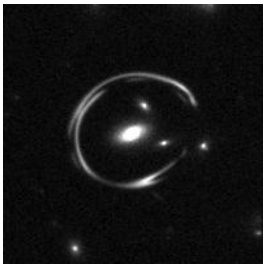
Criterion	Lenses	Non-lenses
assessable 0	$\mu_{eff} \geq 1.6$ & $n_{im} > 0$ & $n_{pix} > 20$	$\mu_{eff} < 1.0$ or $n_{sources} = 0$
assessable 1	$\mu_{eff} \geq 1.6$ & $n_{im} > 0$ & $n_{pix} > 20$	the rest except those with $1 < \mu_{eff} < 1.6$
20-pixel	$\mu_{eff} \geq 4.0$ & $n_{im} > 0$ & $n_{pix} > 20$	$\mu_{eff} < 1.0$ or $n_{sources} = 0$
50-pixel	$\mu_{eff} \geq 4.0$ & $n_{im} > 0$ & $n_{pix} > 50$	$\mu_{eff} < 1.0$ or $n_{sources} = 0$
100-pixel	$\mu_{eff} \geq 4.0$ & $n_{im} > 0$ & $n_{pix} > 100$	$\mu_{eff} < 1.0$ or $n_{sources} = 0$
200-pixel	$\mu_{eff} \geq 4.0$ & $n_{im} > 0$ & $n_{pix} > 200$	$\mu_{eff} < 1.0$ or $n_{sources} = 0$
300-pixel	$\mu_{eff} \geq 4.0$ & $n_{im} > 0$ & $n_{pix} > 300$	$\mu_{eff} < 1.0$ or $n_{sources} = 0$
350-pixel	$\mu_{eff} \geq 4.0$ & $n_{im} > 0$ & $n_{pix} > 350$	$\mu_{eff} < 1.0$ or $n_{sources} = 0$

Table 6.18: Criteria chosen to define the lenses in the multi-band Euclid-like data sets. The criteria from '20-pixel' to '350-pixel' include also the requirement of having a $sb_{contrast} < 10$, in order to avoid a few extreme examples.

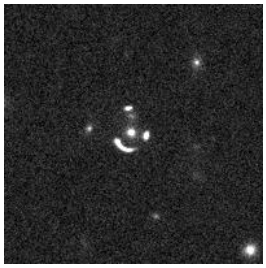
data set	Criterion	Description	Lenses	Non-lenses
Training	20-pixel	15,771 images (66×66) in J, Y, H and VIS bands	8375	7396
Training	50-pixel	14,450 images (66×66) in J, Y, H and VIS bands	7054	7396
Training	100-pixel	12,840 images (66×66) in J, Y, H and VIS bands	5444	7396
Training	200-pixel	10,868 images (66×66) in J, Y, H and VIS bands	3472	7396
Training	300-pixel	9,802 images (66×66) in J, Y, H and VIS bands	2406	7396
Training	350-pixel	9,441 images (66×66) in J, Y, H and VIS bands	2045	7396
Test	assessable 0	35,830 images (66×66) in J, Y, H and VIS bands	30,830	5000
Test	20-pixel	10,000 images (66×66) in J, Y, H and VIS bands	5000	5000
Test	50-pixel	9,183 images (66×66) in J, Y, H and VIS bands	4183	5000
Test	100-pixel	8,235 images (66×66) in J, Y, H and VIS bands	3235	5000
Test	200-pixel	7,026 images (66×66) in J, Y, H and VIS bands	2026	5000
Test	300-pixel	6,410 images (66×66) in J, Y, H and VIS bands	1410	5000
Test	350-pixel	6,174 images (66×66) in J, Y, H and VIS bands	1174	5000
Challenge	assessable 1	88,910 images (66×66) in J, Y, H and VIS bands	49,596	39,314

Table 6.19: Description of the data sets for the multi-band Euclid-like simulations.

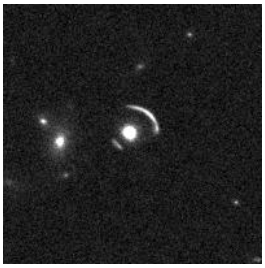
The greater realism of the simulations we discussed above also brings a greater variety of scenarios, resulting more difficult to know by eye what we are seeing, even after checking the catalog with the properties of the candidates. Some examples of the candidates within these simulations are shown in Fig 6.26. In these images we can see evident lenses with high μ_{eff} (Fig. 6.26a), clear lenses with lower μ_{eff} (Figures 6.26b to 6.26d), unclear spiral-like candidates with really high μ_{eff} (Fig. 6.26e), off center lenses (Fig. 6.26f), high-μ_{eff} candidates with zero images of the source (Figures 6.26g to 6.26i), and low-μ_{eff} candidates with many images of the source (Figures 6.26j to 6.26l). These cases with many images but low μ_{eff} can be due to lensing by a galaxy in a cluster. Some of the images might be outside the field of view shown in the cutout, and since μ_{eff} is calculated over the pixels in the image, its value is lower than it should. Most of the candidates with zero images, but high effective magnification, are probably cases where the source was too dim, despite μ_{eff}, to be above the threshold. Other interesting examples, like that in Fig 6.26e with only 1 image of the source but $\mu_{eff} = 44.74$, can be caused by the merging of several images into one.



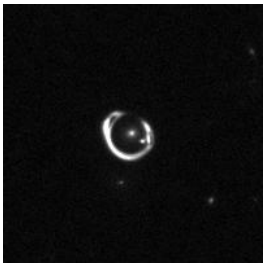
(a) ID = 21
(μ_{eff} = 17.94, n_{im} = 2)



(b) ID = 21886
(μ_{eff} = 5.23, n_{im} = 3)



(c) ID = 79
(μ_{eff} = 3.01, n_{im} = 2)

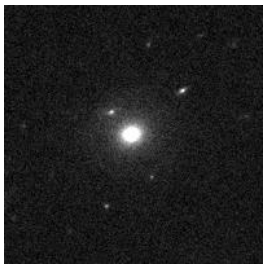


(d) ID = 95
(μ_{eff} = 3.17, n_{im} = 1)

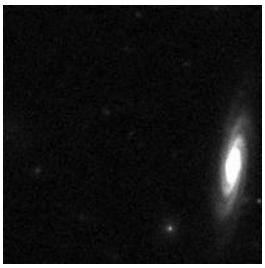
(e) ID = 847
(μ_{eff} = 44.74, n_{im} = 1)

(f) ID = 89467
(μ_{eff} = 13.98, n_{im} = 1)

(g) ID = 22275
(μ_{eff} = 13.45, n_{im} = 0)

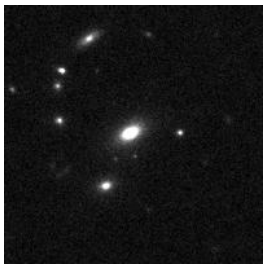


(h) ID = 37369
(μ_{eff} = 14.40, n_{im} = 0)



(i) ID = 88041
(μ_{eff} = 19.02, n_{im} = 0)

(j) ID = 38
(μ_{eff} = 0.46, n_{im} = 4)



(k) ID = 312
(μ_{eff} = 1.45, n_{im} = 4)

(l) ID = 92666
(μ_{eff} = 1.95, n_{im} = 7)

Figure 6.26: Some examples in the visible band of the multi-band Euclid-like data.

6.3.2 Results for multi-band Euclid-like data

We have put our network through six learning processes using the training data sets determined in Table 6.19. 10% of the examples in every training data set were used for validation. Data from every example were normalized to the maximum value in each band prior to be input to the network. And the network architecture used is the same as described in Sec. 5.3, with the exception that input data are now four-band. The early-stopping tool is again used here. The 20-pixel and 50-pixel models were achieved after training the network over ~ 60 epochs, while the rest of models required ~ 40 epochs. The number of trainable parameters of the network per each layer for all the trainings carried out are shown in Table 6.20.

Layer	Number of weights
1st convolutional layer	4,640
2nd convolutional layer	51,264
3rd convolutional layer	32,896
4th convolutional layer	147,584
1st fully connected layer	2,097,216
2nd fully connected layer	65
Full network	2,333,665

Table 6.20: Number of weights of the CNN trained with multi-band Euclid-like simulations.

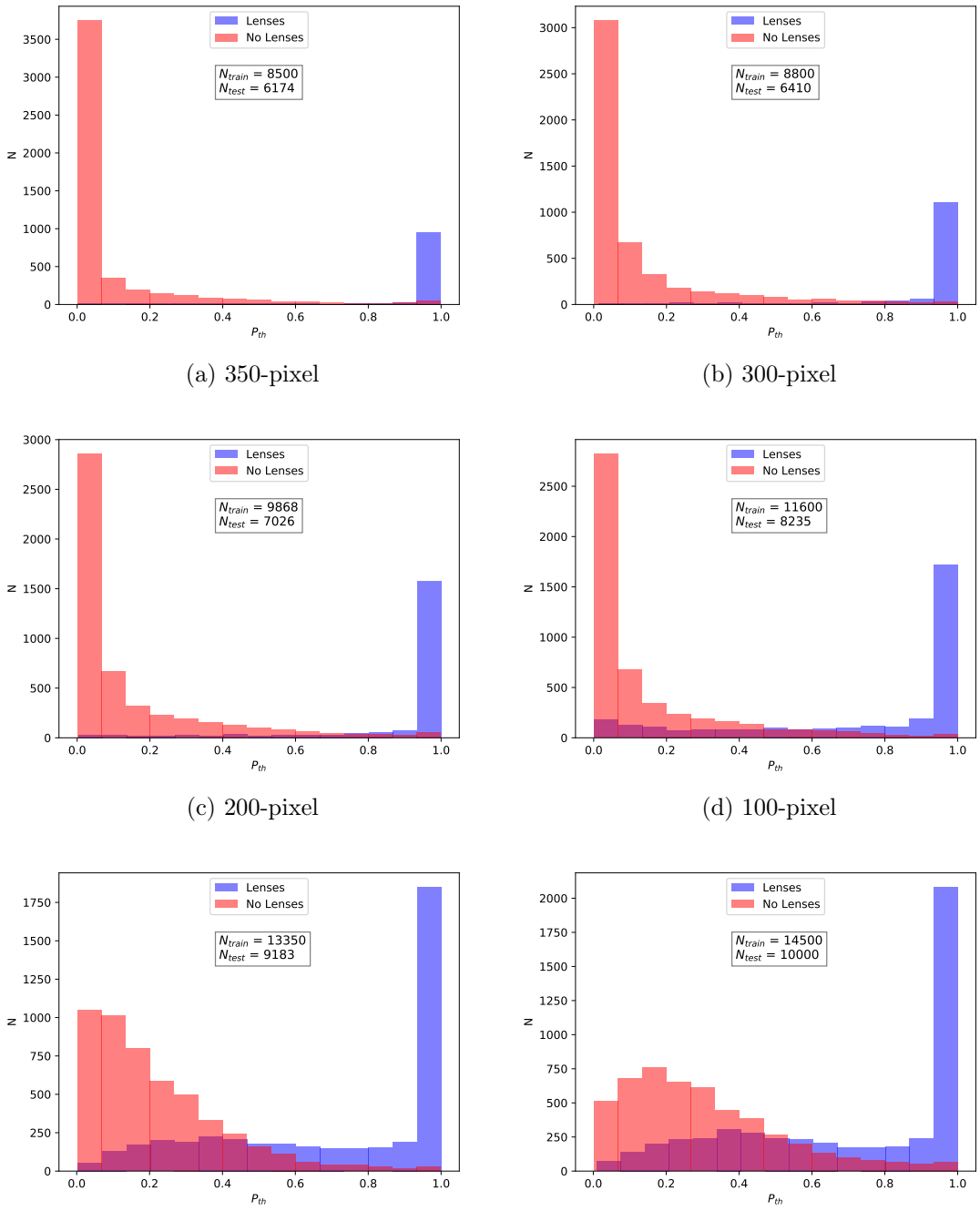
In first place, we applied each model derived with the network to the corresponding test data set, built following the same criteria as the ones used in the training data sets. This is, the model trained with the 20-pixel training data set was applied to the 20-pixel test data set, and so on. In this way, we can check out the capacity of the different models to identify lenses defined following the same rules as those used to train the models. The resulting probability distributions are shown in Fig. 6.27. As can be seen in them, the more aggressive is the criterion requested to define the lenses, the better is the separation achieved between both classes. However, the ability of the models to classify the non-lens class worsens more than for the lens class when we go down to the 50-pixel and 20-pixel criteria. According to all probability distributions, we can easily extract a pure enough sample of lenses by demanding a high cut in the probability threshold. The confusion matrices from Table 6.21 show how the TPR is kept above 90% for the three stricter models. Then it starts going down until it reaches its minimum value for the 20-pixel model ($TPR = 44.5\%$). While the accuracy decreases as the criteria to define the lenses are softer, the precision increases until it reaches its maximum value for the 50-pixel model

($Pre = 87.9\%$). The ROC curves in Fig. 6.28 show clearly how the ability of the models in distinguishing between classes gets better as the criteria to define the lenses are more stringent. As these curves demonstrate, the 350-pixel model is the best classifier between lenses and non-lenses. Nevertheless, in this occassion models are compared by calculating the maximum value of F_β reached for any P_{th}. A value of $\beta = 0.03$ was assummed so, according to Eq. (5.16), those models with the least contaminated lens classifications are expected to score higher F_β values. The curves of F_β as a function of the probability threshold for all the models considered are depicted in Fig. 6.29. The maximum value of F_β was calculated for every model, and we inferred that the best model is that trained with the 50-pixel criterion (max $F_\beta = 0.9907$). This is consistent with the information we already had, since the 50-pixel model supplies us with the most precise sample of lenses.

In second place, we have applied all the models derived with the network to the test data set built with the 'assessable 0' criteria, in which non-lenses are defined as those examples with $\mu_{eff} < 1.0$ or no sources added. Given that all models are being tested on the same data set with a criterion almost identical to the 'assessable 1', the results achieved this time can give us a more accurate idea of the perfomance our models will have over the challenge data set. In the probability distributions from Fig. 6.30, one can see that the 350-pixel model houses the largest number of lenses missclassified as non-lenses in the 0-valued bin, and the lower amount of true lenses correctly classified as lenses in the 1-valued bin. It is also the model that achieves the best classification of the non-lenses, scoring almost all of them with the lowest probabilities. As we consider models trained with looser criteria to define the lenses, this scenario evolves auspiciously for the lenses. The wrongly classified lenses tend to move to the right, emptying the 0-valued bin and increasing their number in the 1-valued bin. However, the non-lenses suffer the opposite effect, ceasing to be so well classified as they were with the stricter models. This behavior is also reflected in the confusion matrices from Table 6.22. Unlike what happened when we applied the models to the test data sets built with the same criteria as those used to train the models, the TPR now increases from 20.0%, in the strictest model, to 42.5%, in the softer model. The same happens to the accuracy, which rises from 30.5%,, for the 350-pixel model, to 48.2%, for the 20-pixel model. And here the precision remains about constant, between $\sim 94\%$ and $\sim 96\%$ for all models, achieving its highest value for the 50-pixel model (96.4%). Looking at it all, although the 20-pixel model accomplishes the greatest accuracy and TPR, the 50-pixel model is the more precise, getting to reduce by half, from 16.3% to 8.0%, the amount of false positive examples of the former. This evidence supports the fact that the 50-pixel model not only is once again the best model, attending to the F_β score (see Fig. 6.32), but it has also managed to establish as the best model according to the ROC curves (see Fig. 6.31).

In the end, we have applied all the models to the challenge data set, in which, in contradistinction to the previous test data set, all those candidates that do not meet the requirements to be lenses are considered non-lenses. Thus, the amount of non-lenses increases meaningly. The behavior of the models that we observed with the previous data set reappears here. Therefore, it was a helpful test in order to forecast the performance of our models, and to decide which one would yield better results in the challenge. The probability distributions are shown in Fig. 6.33, and the confusion matrices can be read in Table 6.23. According as we lower the level of demand to define the lenses, the classification of lenses improves but that of non-lenses gets worse. We notice an increase of the TPR from 22.5% to 44.5%, and the accuracy moves up from 54.7% to 61.0%. The precision remains steady, between $\sim$ 83% and $\sim$ 86%, for all models except the 20-pixel model, in which it goes down to 75.5%. The 50-pixel model achieves the best accuracy, and the second largest values for the precision and TPR, when taking $P_{th} = 0.5$. In Table 6.24 we see that it does reach greater precisions than the rest of the models, when taking higher cuts for the probability scores: 92.6% with $P_{th} = 0.7$, and 97.6% with $P_{th} = 0.9$. Finally, as it follows from Figures 6.34 and 6.35, the 50-pixel model is the one that exhibits a better overall performance in distinguishing between lenses and non-lenses (AUC = 0.700), and provides a purer sample of lenses (maximum $F_\beta = 0.9859$). We show in Fig. 6.36 the evolution of the precision, from all the models considered, as a function of some of the characteristics of the lenses, such as μ_{eff}, n_{pix}, the redshifts of source and lens, or the mass of the main lens in the simulation. In this plots we can see that the 50-pixel model outperforms the rest of models, except for lensing systems with sources beyond $z \approx 5.5$.

The 50-pixel model accomplishes 8837 true positives, i.e., true lenses identified, with $P_{th} > 0.9$, and 3147 true positives with $P_{th} = 1.0$. We show in Fig. 6.37 a random sample of 35 true positive examples with $P_{th} = 1.0$. As it can be easily seen, all of them are evident lenses with multiple arcs and/or ring-shaped images. However, this model also presents 220 false positives, i.e., non-lenses wrongly identified as lenses, with $P_{th} > 0.9$. In Fig. 6.38 we show a random sample of 35 false positive examples with $P_{th} > 0.99$. Among them we can see, for instance, a significant amount of spiral-like candidates, but also 12 lenses recognizable by eye. These are: 9361, 24322, 27162, 32049, 32429, 43513, 54910, 58551, 76698, 79016, 89898, and 99425. Other 3 examples (29270, 40512, and 64317) could also be lenses, but we are less confident about them. The fact that a third part of the examples, in this random sample of false positives, are evident lenses tells us that our models are capable of identifying lenses below the criteria defined in the challenge to decide whether examples are or not lenses. These are good news, since in blind searches for lenses in large-scale surveys we do not have this information, and all the positive identifications must be examined.



(a) 350-pixel (b) 300-pixel

(c) 200-pixel (d) 100-pixel

(e) 50-pixel (f) 20-pixel

Figure 6.27: Probability distributions obtained with the models derived using the multi-band Euclid-like training data sets defined in Table 6.19, when applied to test samples built following the same criteria as the training ones to define the lenses. Thus, Fig. 6.27a shows the results of applying a model trained with the 350-pixel training data set to the 350-pixel test data set. The rest of plots are read the same way. Blue bins represent those examples in the test sample that are lenses while the red bins represent the non-lenses.

True \ Predicted	Non-lens	Lens
Non-lens	4760 (95.2%)	240 (4.8%)
Lens	105 (8.9%)	1069 (91.1%)

(a) 350-pixel ($Acc = 94.4\%$. $Pre = 81.7\%$).

True \ Predicted	Non-lens	Lens
Non-lens	4677 (93.5%)	323 (6.5%)
Lens	104 (7.4%)	1306 (92.6%)

(b) 300-pixel ($Acc = 93.3\%$. $Pre = 80.2\%$).

True \ Predicted	Non-lens	Lens
Non-lens	4612 (92.2%)	388 (7.8%)
Lens	177 (8.7%)	1849 (91.3%)

(c) 200-pixel ($Acc = 92.0\%$. $Pre = 82.7\%$).

True \ Predicted	Non-lens	Lens
Non-lens	4613 (92.3%)	387 (7.7%)
Lens	764 (23.6%)	2471 (76.4%)

(d) 100-pixel ($Acc = 86.0\%$. $Pre = 86.5\%$).

True \ Predicted	Non-lens	Lens
Non-lens	4598 (92.0%)	402 (8.0%)
Lens	1263 (30.2%)	2920 (69.8%)

(e) 50-pixel ($Acc = 81.9\%$. $Pre = 87.9\%$).

True \ Predicted	Non-lens	Lens
Non-lens	4187 (81.8%)	813 (18.2%)
Lens	1571 (55.5%)	3429 (44.5%)

(f) 20-pixel ($Acc = 76.2\%$. $Pre = 80.8\%$).

Table 6.21: Confusion matrices obtained with the models derived using the multi-band Euclid-like training data sets defined in Table 6.19, when applied to test samples built following the same criteria as the training ones to define the lenses. Thus, Table 6.21a shows the results of applying a model trained with the 350-pixel training data set to the 350-pixel test data set. The rest of matrices are read the same way. We chose $P_{thr} = 0.5$ as the reference probability threshold to decide whether or not an example is predicted as a lens. The accuracy (Acc from Eq. 5.14) and precision (Pre from Eq. 5.15) of the different models are also calculated based on this threshold.

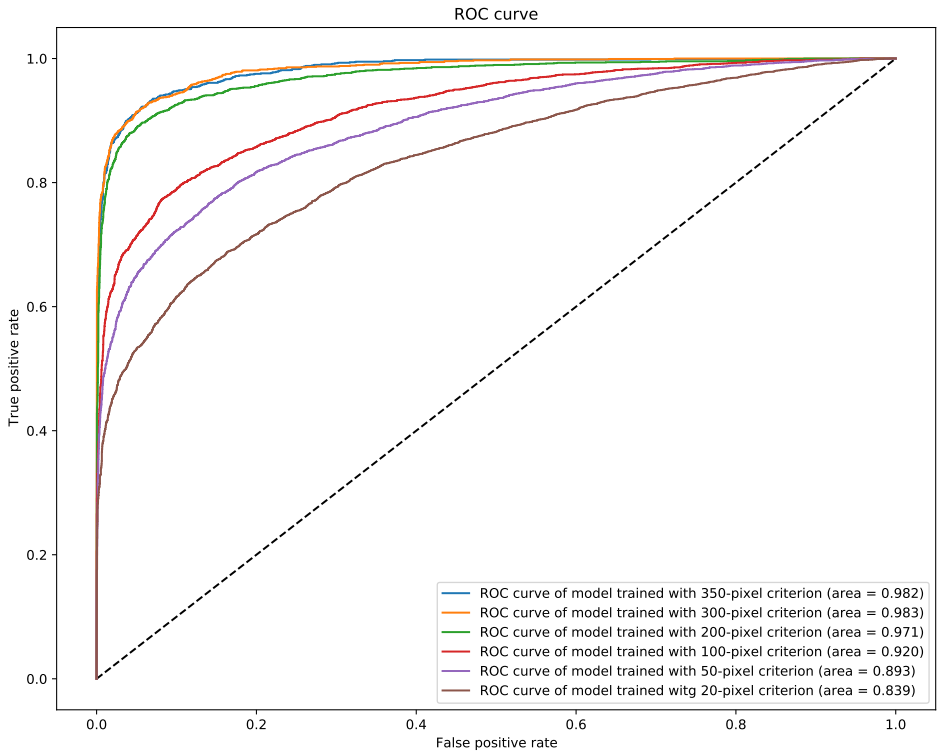


Figure 6.28: ROC curves of models derived using the multi-band Euclid-like training data sets defined in Table 6.19, when applied to test samples built following the same criteria as the training ones to define the lenses. Thus, for example, the red line corresponds to a model trained with the 100-pixel training data set and applied to the 100-pixel test data set. The rest of curves are read the same way.

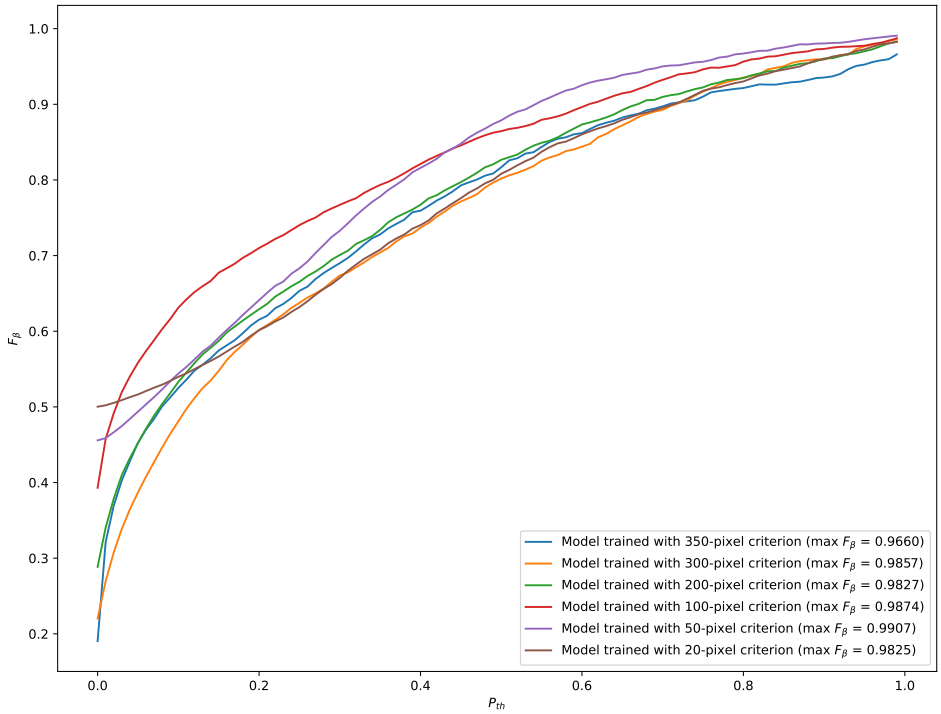
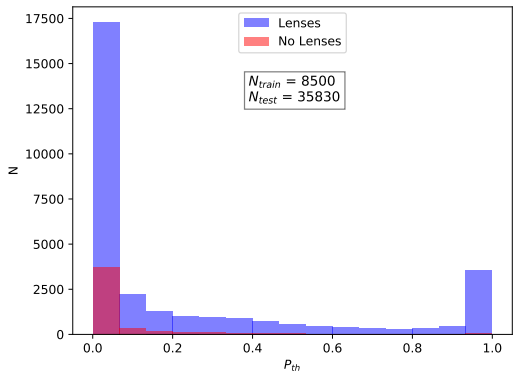
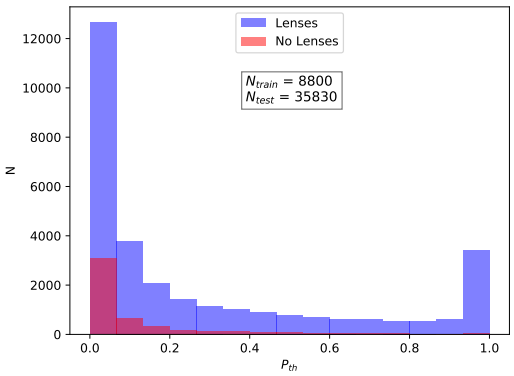


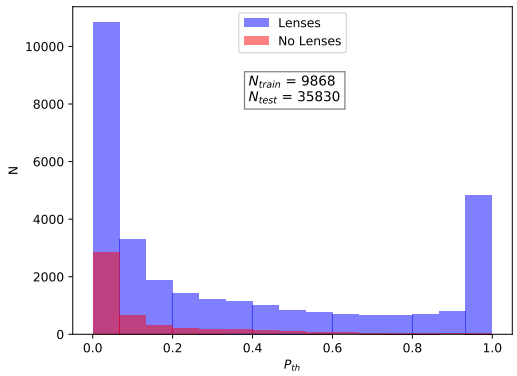
Figure 6.29: F_β curves, as a function of the probability threshold, of models derived using the multi-band Euclid-like training data sets defined in Table 6.19, when applied to test samples built following the same criteria as the training ones to define the lenses. Thus, for example, the red line corresponds to a model trained with the 100-pixel training data set and applied to the 100-pixel test data set. The rest of curves are read the same way. F_β values were calculated using Eq. (5.16) assumming $\beta = 0.03$. For each model the maximum F_β reached for any P_{thr} is estimated.



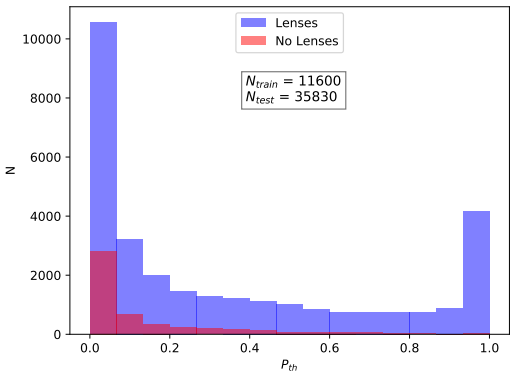
(a) 350-pixel



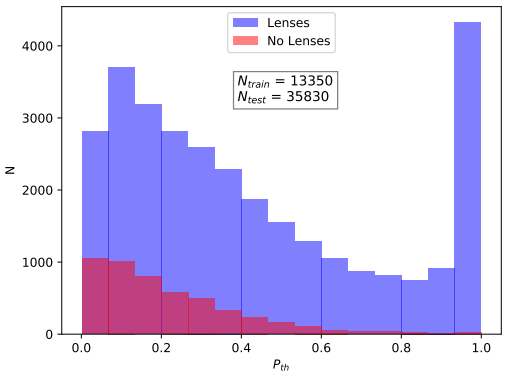
(b) 300-pixel



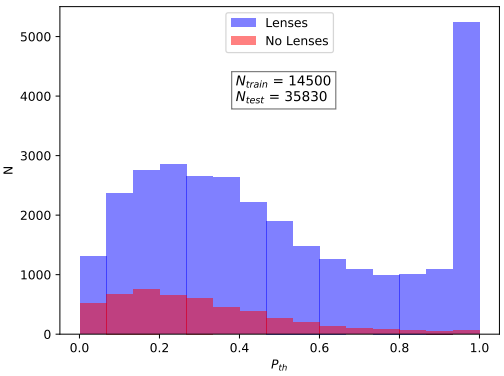
(c) 200-pixel



(d) 100-pixel



(e) 50-pixel



(f) 20-pixel

Figure 6.30: Probability distributions obtained with the models derived using the multi-band Euclid-like training data sets defined in Table 6.19, when applied to test samples built following the 'assessable 0' criteria to define the lenses. Thus, Fig. 6.30a shows the results of applying a model trained with the 350-pixel training data set to the 'assessable 0' test data set. The rest of plots are read the same way. Blue bins represent those examples in the test sample that are lenses while the red bins represent the non-lenses.

		Predicted	
		Non-lens	Lens
True	Non-lens	4760 (95.2%)	240 (4.8%)
	Lens	24661 (80.0%)	6169 (20.0%)

(a) 350-pixel ($Acc = 30.5\%$. $Pre = 96.3\%$).

		Predicted	
		Non-lens	Lens
True	Non-lens	4677 (93.5%)	323 (6.5%)
	Lens	23422 (76.0%)	7408 (24.0%)

(b) 300-pixel ($Acc = 33.7\%$. $Pre = 95.8\%$).

		Predicted	
		Non-lens	Lens
True	Non-lens	4612 (92.2%)	388 (7.8%)
	Lens	21311 (69.1%)	9519 (30.9%)

(c) 200-pixel ($Acc = 39.4\%$. $Pre = 96.1\%$).

		Predicted	
		Non-lens	Lens
True	Non-lens	4613 (92.3%)	387 (7.7%)
	Lens	21427 (69.5%)	9403 (30.5%)

(d) 100-pixel ($Acc = 39.1\%$. $Pre = 96.0\%$).

		Predicted	
		Non-lens	Lens
True	Non-lens	4598 (92.0%)	402 (8.0%)
	Lens	20114 (65.2%)	10716 (34.8%)

(e) 50-pixel ($Acc = 42.7\%$. $Pre = 96.4\%$).

		Predicted	
		Non-lens	Lens
True	Non-lens	4187 (83.7%)	813 (16.3%)
	Lens	17730 (57.5%)	13100 (42.5%)

(f) 20-pixel ($Acc = 48.2\%$. $Pre = 94.2\%$).

Table 6.22: Confusion matrices obtained with the models derived using the multi-band Euclid-like training data sets defined in Table 6.19, when applied to test samples built following the 'assessable 0' criteria to define the lenses. Thus, Table 6.22a shows the results of applying a model trained with the 350-pixel training data set to the 'assessable 0' test data set. The rest of matrices are read the same way. We chose $P_{thr} = 0.5$ as the reference probability threshold to decide whether or not an example is predicted as a lens. The accuracy (Acc from Eq. 5.14) and precision (Pre from Eq. 5.15) of the different models are also calculated based on this threshold.

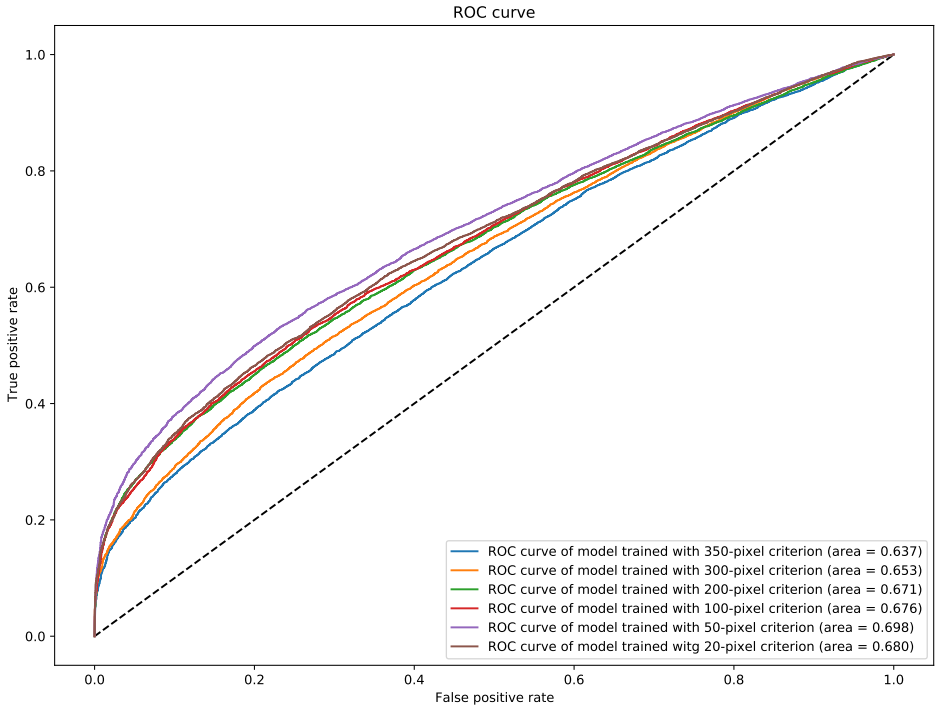


Figure 6.31: ROC curves of models derived using the multi-band Euclid-like training data sets defined in Table 6.19, when applied to test samples built following the 'assessable 0' criteria to define the lenses. Thus, for example, the red line corresponds to a model trained with the 100-pixel training data set and applied to the 'assessable 0' test data set. The rest of curves are read the same way.

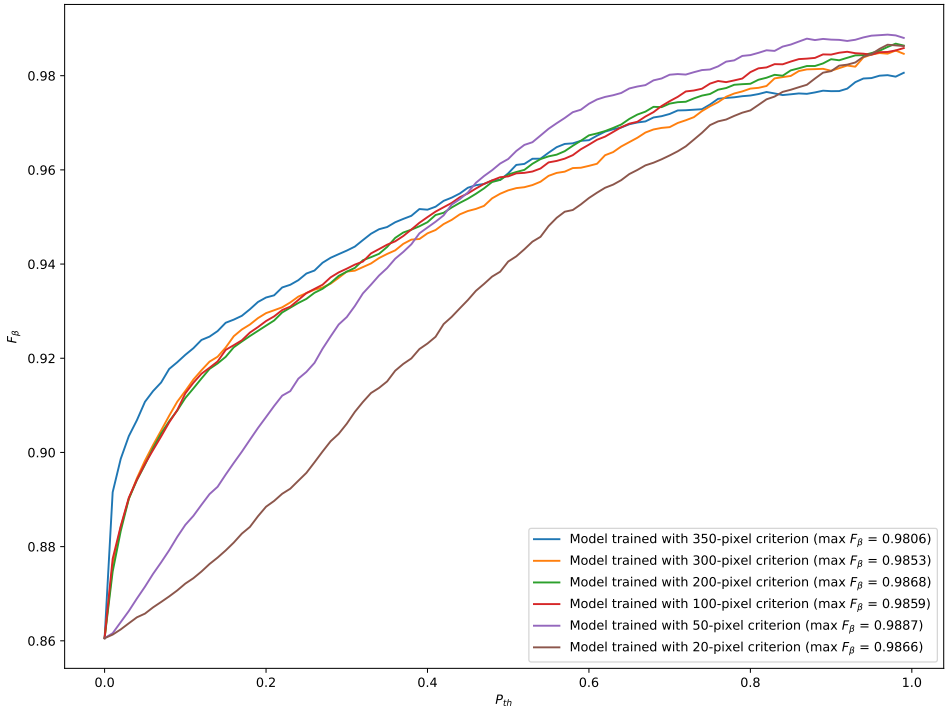
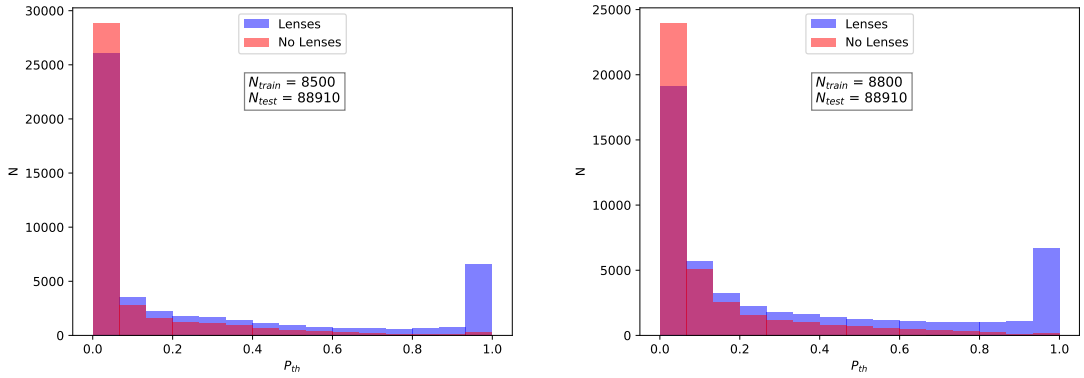
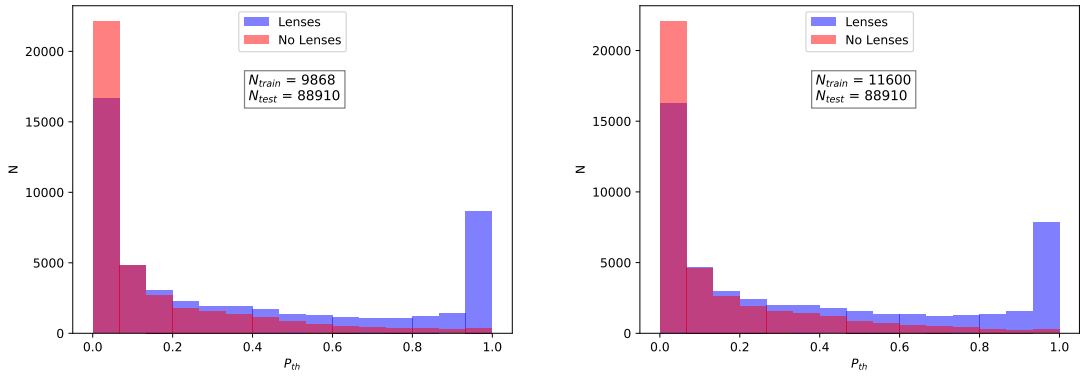


Figure 6.32: F_β curves, as a function of the probability threshold, of models derived using the multi-band Euclid-like training data sets defined in Table 6.19, when applied to test samples built following the 'assessable 0' criteria to define the lenses. Thus, for example, the red line corresponds to a model trained with the 100-pixel training data set and applied to the 'assessable 0' test data set. The rest of curves are read the same way. F_β values were calculated using Eq. (5.16) assumming $\beta = 0.03$. For each model the maximum F_β reached for any P_{thr} is estimated.



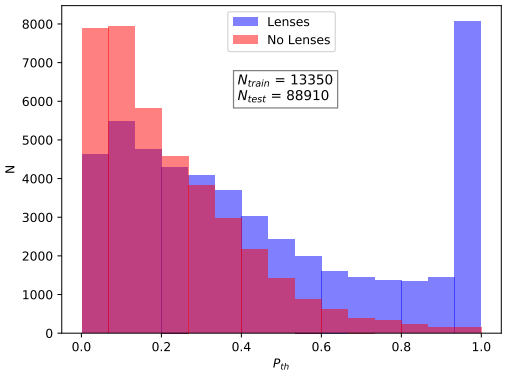
(a) 350-pixel

(b) 300-pixel

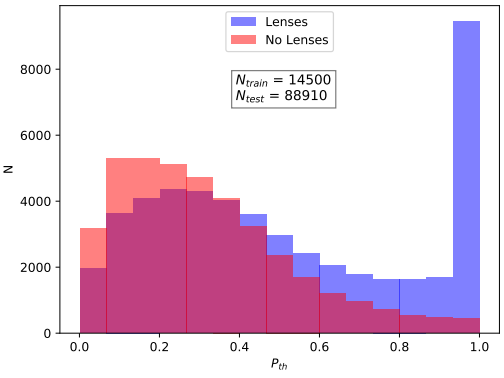


(c) 200-pixel

(d) 100-pixel



(e) 50-pixel



(f) 20-pixel

Figure 6.33: Probability distributions obtained with the models derived using the multi-band Euclid-like training data sets defined in Table 6.19, when applied to the challenge data set. Blue bins represent those examples in the test sample that are lenses while the red bins represent the non-lenses.

		Predicted	
		Non-lens	Lens
True	Non-lens	37475 (95.3%)	1839 (4.7%)
	Lens	38446 (77.5%)	11150 (22.5%)

(a) 350-pixel ($Acc = 54.7\%$. $Pre = 85.8\%$).

		Predicted	
		Non-lens	Lens
True	Non-lens	36644 (93.2%)	2670 (6.8%)
	Lens	35918 (72.4%)	13678 (27.6%)

(b) 300-pixel ($Acc = 56.6\%$. $Pre = 83.7\%$).

		Predicted	
		Non-lens	Lens
True	Non-lens	35981 (91.5%)	3333 (8.5%)
	Lens	33025 (66.6%)	16571 (33.4%)

(c) 200-pixel ($Acc = 59.1\%$. $Pre = 83.3\%$).

		Predicted	
		Non-lens	Lens
True	Non-lens	35911 (91.3%)	3403 (8.7%)
	Lens	32900 (66.3%)	16696 (33.7%)

(d) 100-pixel ($Acc = 59.2\%$. $Pre = 83.1\%$).

		Predicted	
		Non-lens	Lens
True	Non-lens	35960 (91.5%)	3354 (8.5%)
	Lens	31158 (62.8%)	18438 (37.2%)

(e) 50-pixel ($Acc = 61.2\%$. $Pre = 84.6\%$).

		Predicted	
		Non-lens	Lens
True	Non-lens	32156 (82.1%)	7158 (17.9%)
	Lens	27514 (55.5%)	22082 (44.5%)

(f) 20-pixel ($Acc = 61.0\%$. $Pre = 75.5\%$).

Table 6.23: Confusion matrices obtained with the models derived using the multi-band Euclid-like training data sets defined in Table 6.19, when applied to the challenge data set. We chose $P_{thr} = 0.5$ as the reference probability threshold to decide whether or not an example is predicted as a lens. The accuracy (Acc from Eq. 5.14) and precision (Pre from Eq. 5.15) of the different models are also calculated based on this threshold.

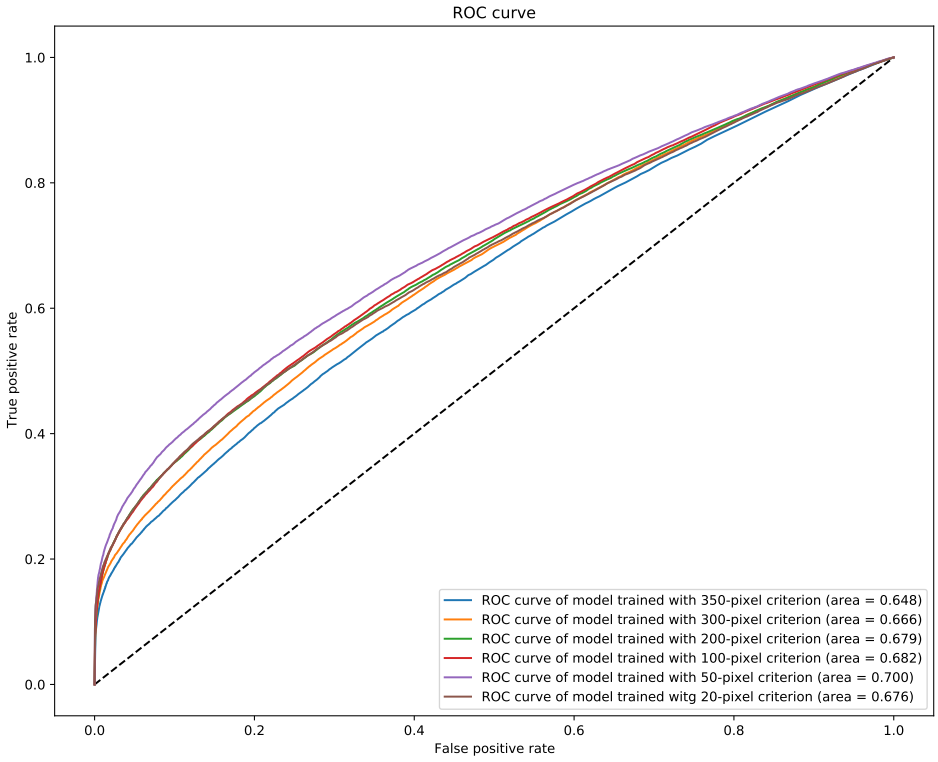


Figure 6.34: ROC curves of models derived using the multi-band Euclid-like training data sets defined in Table 6.19, when applied to the challenge data set.

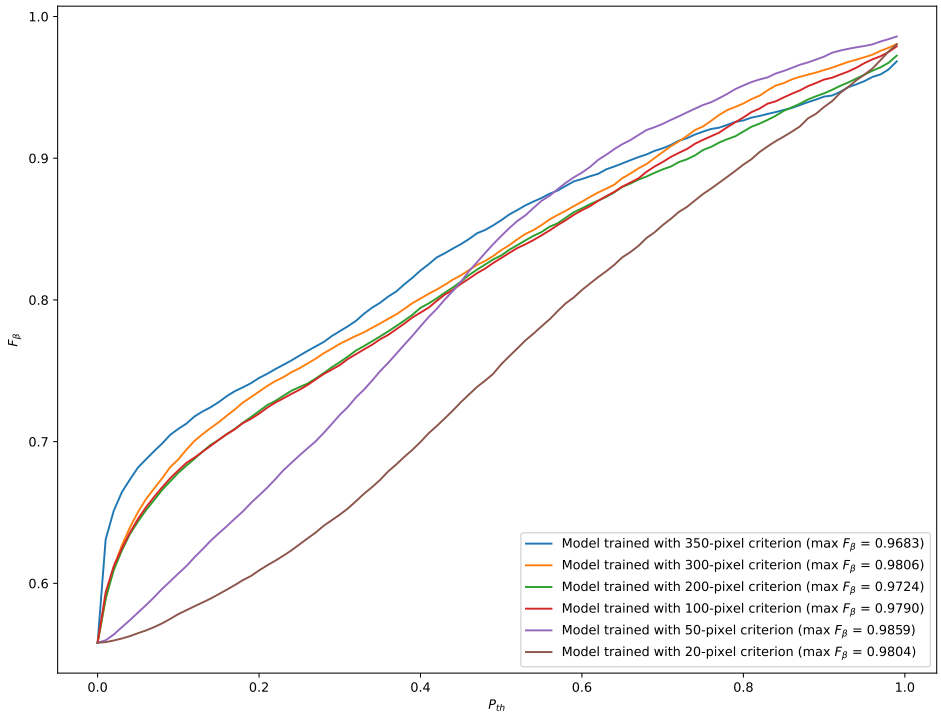
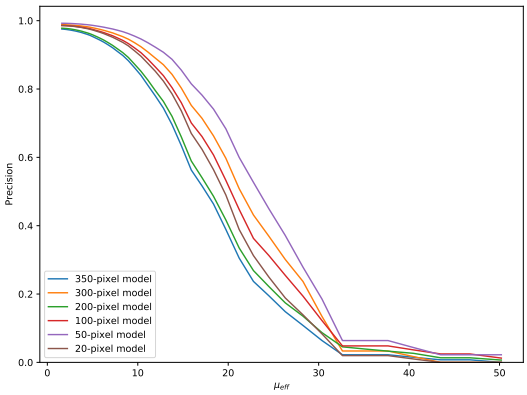


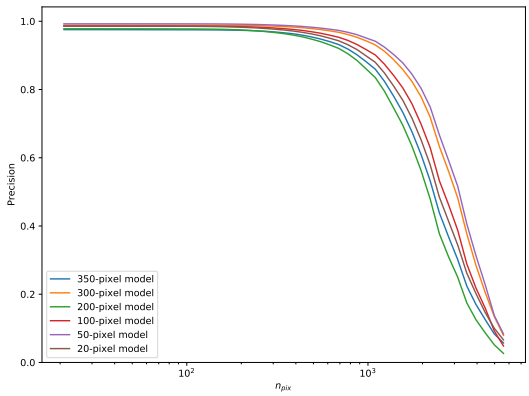
Figure 6.35: F_β curves, as a function of the probability threshold, of models derived using the multi-band Euclid-like training data sets defined in Table 6.19, when applied to the challenge data set. F_β values were calculated using Eq. (5.16) assumming $\beta = 0.03$. For each model the maximum F_β reached for any P_{thr} is estimated.

Test data set	P_{thr}	TPR	Acc	Pre
model trained with 350-pixel training data set	0.5	22.5%	54.7%	85.8%
	0.7	18.0%	53.3%	91.0%
	0.9	14.2%	51.7%	94.8%
model trained with 300-pixel training data set	0.5	27.6%	56.6%	83.7%
	0.7	20.8%	54.6%	90.7%
	0.9	14.6%	52.1%	96.7%
model trained with 200-pixel training data set	0.5	33.4%	59.1%	83.3%
	0.7	26.1%	57.0%	89.4%
	0.9	19.0%	54.3%	94.9%
model trained with 100-pixel training data set	0.5	33.7%	59.2%	83.1%
	0.7	25.4%	56.8%	89.9%
	0.9	17.4%	53.5%	95.9%
model trained with 50-pixel training data set	0.5	37.2%	61.2%	84.6%
	0.7	26.0%	57.6%	92.6%
	0.9	17.8%	53.9%	97.6%
model trained with 20-pixel training data set	0.5	44.5%	61.0%	75.5%
	0.7	30.9%	58.5%	85.4%
	0.9	20.9%	55.1%	93.9%

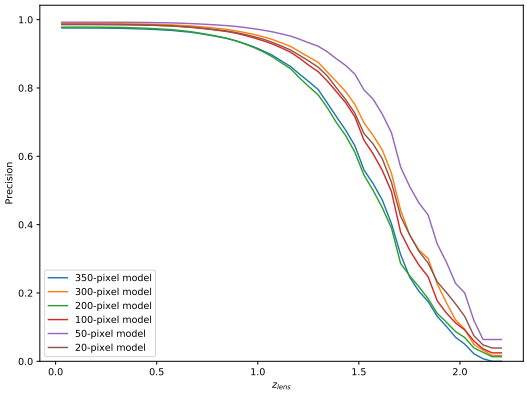
Table 6.24: Recall (TPR), accuracy (Acc) and precision (Pre) values of all the models applied to multi-band Euclid-like challenge data set considering three different P_{thr} cuts.



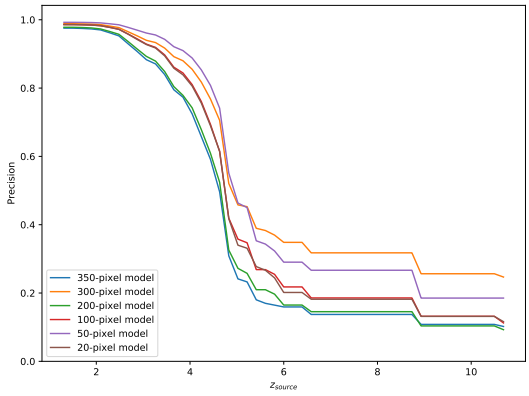
(a) μ_{eff}



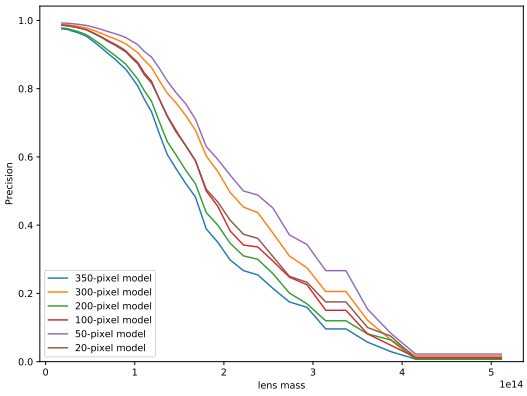
(b) n_{pix}



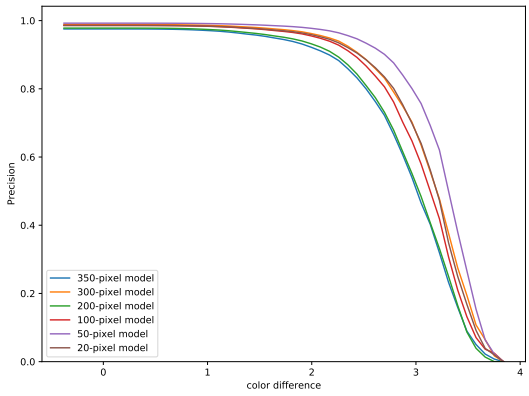
(c) z_{lens}



(d) z_{source}



(e) Lens mass



(f) Color difference between lens and source

Figure 6.36: Precision as a function of several parameters of the multi-band Euclid-like simulated data, for all the models considered. These parameters are the effective magnification of the source (Fig. 6.36a), the number of pixels in the lensed source image abova 1σ (Fig. 6.36b), the redshift of the lens (Fig. 6.36c), the redshift of the source (Fig. 6.36d), the halo mass of the main lens (Fig. 6.36e), and the difference in color between lens and source (Fig. 6.36f).

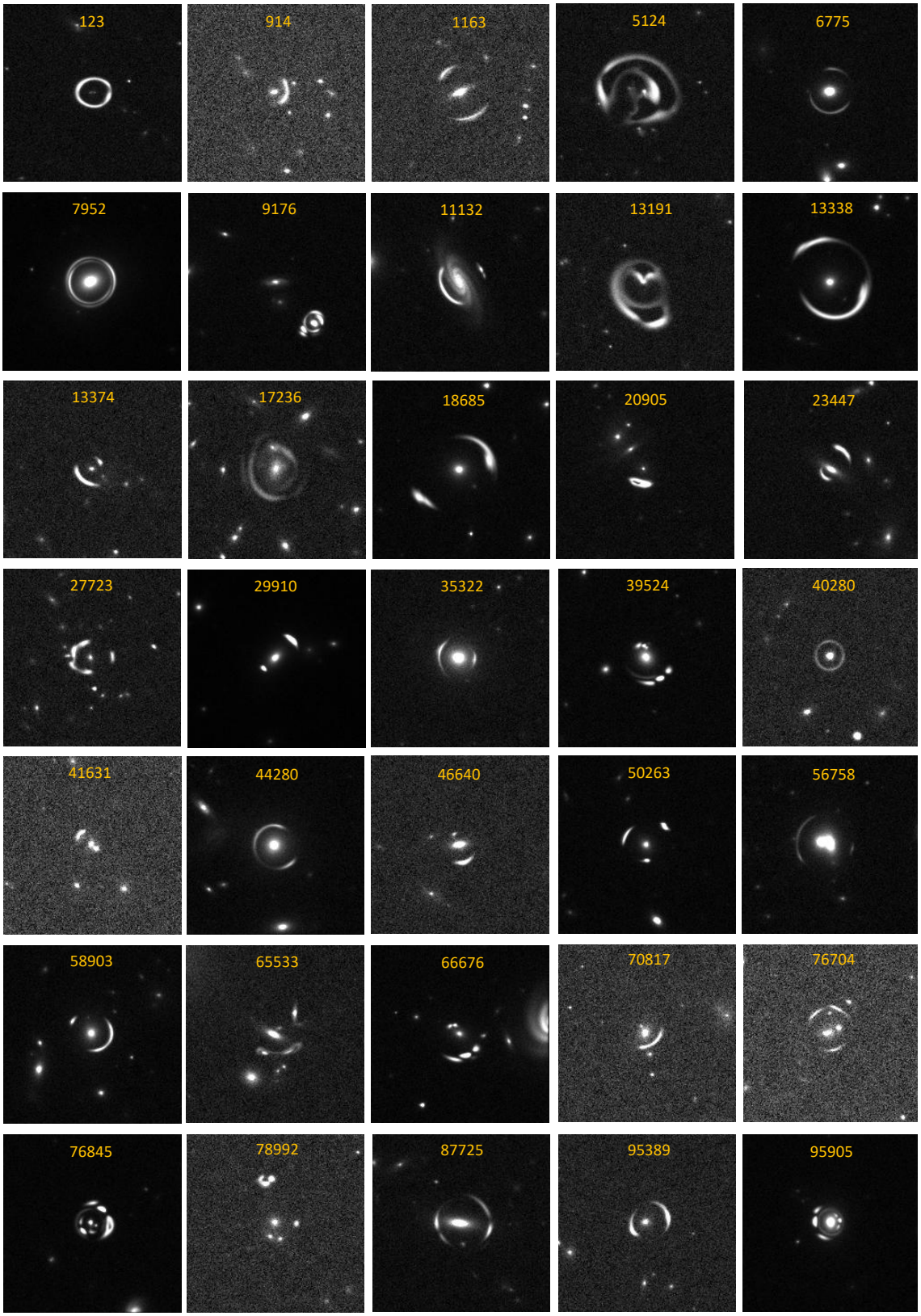


Figure 6.37: Images in the visible band of some high scoring true positive examples from the multi-band Euclid-like challenge data set, i.e., lenses predicted correctly as lenses by our 50-pixel model with $P_{th} = 1.0$.

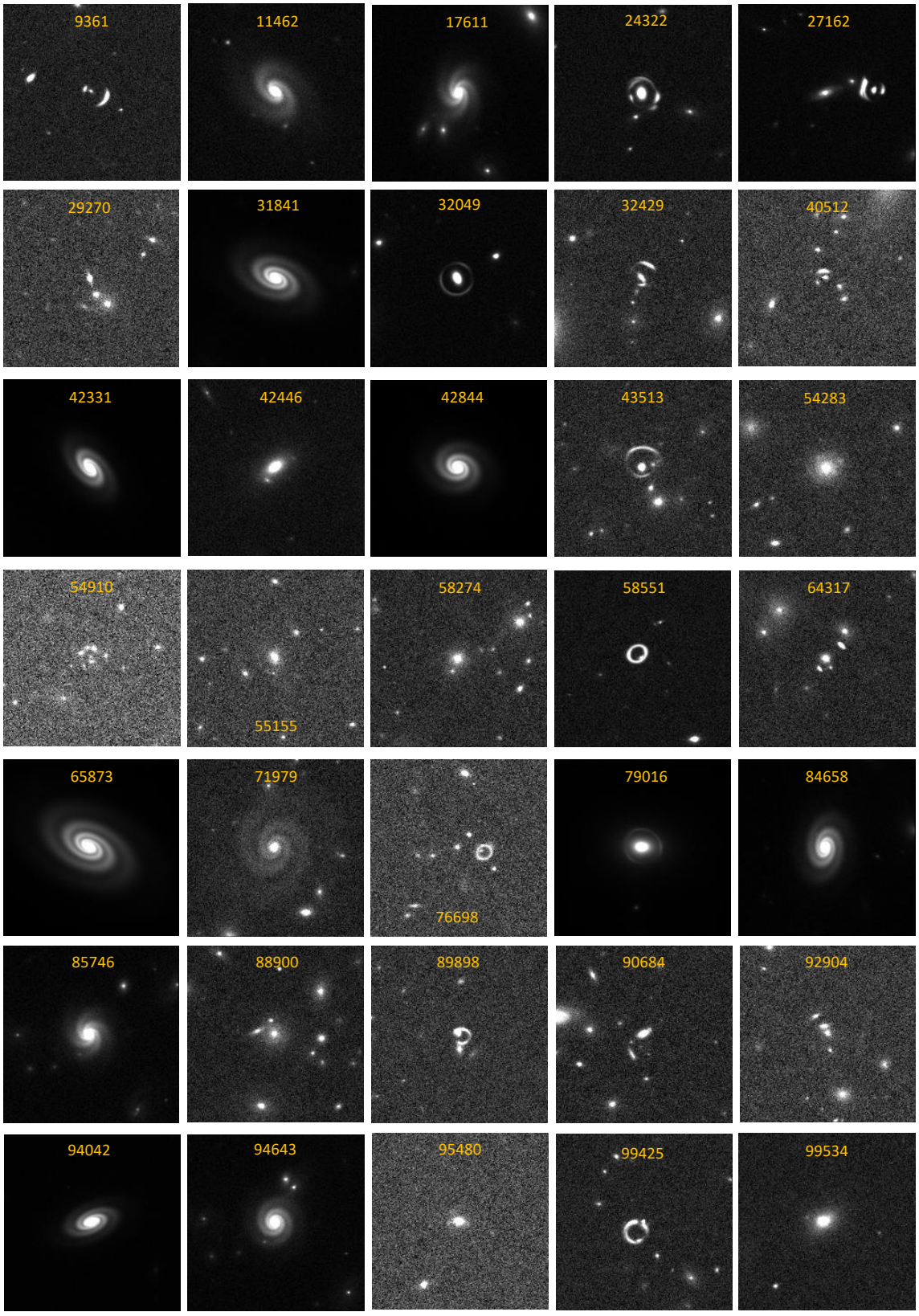


Figure 6.38: Images in the visible band of some high scoring false positive examples from the multi-band Euclid-like challenge data set, i.e., non-lenses predicted as lenses by our 50-pixel model with $P_{th} > 0.99$.

6.4 Conclusions

In this chapter the performance in finding strong gravitational lenses of several models based on the same convolutional neural network is presented. We considered, both for training and testing, data mimicking or proceeding from different astronomical surveys. In first place, several trials using KiDS-like simulated data were carried out in order to lay the ground. We found that a normalization of the input data for every example to the maximum value in each band considered, reducing the amount of dropout used and early-stopping when the performance on a validation sample stops improving yields the best seen results. In the specific case of KiDS-like data, random artificial masking appears for some pixels in 13% of the examples, which we think were included to mimic noisy data or a bad response from the detector. Since we deem these masks are excessive, hidding in some cases completely the lens or non-lens features in the images, and we achieve better results without them, we decided to remove these examples from the training and test data sets.

The standard figure of merit used to compare the performance of classification models is the area under the ROC curve (AUC). The best model trained using KiDS-like data achieves an AUC = 0.955 when tested on the full test data set, and an AUC = 0.982 when the examples containing artificial masks are removed (see Fig. 6.16). A 15% of the examples in these test data sets used actual images taken from a preliminary sample of bright galaxies observed with the KiDS survey. When our model is applied to these concrete examples, the AUC falls to 0.832, but with a model trained and tested using exclusively these examples we achieve an AUC = 0.888 (see Fig. 6.17). This is a clear example of the efficiency differences that may arise when models trained using simulations are tested on real data. And it is thus important that the simulations are as realistic as possible. Our best model was later tested on images from a sample of SDSS and DES strong lens candidates, achieving precisions of 65% and 83%, respectively. Images of some of these candidates, together with the probabilities given by our model, are shown in Figures 6.18, 6.19, and 6.20. It has also been applied to the 40 potential lens candidates found in Sec. 2.7, among which 7 are predicted as lenses, and to a set of potential lenses identified thanks to the Galaxy Zoo 2 project, an subsequently checked by eye. 86% of these last candidates were predicted to be lenses by our model (see Fig. 6.21).

On the other hand, the best model trained with single-band Euclid-like data, which was obtained by using slightly stricter criteria to define the lenses than those used to evaluate the model on the test data set, delivers an AUC = 0.921 (see Fig. 6.25). As it can be seen, despite the fact that Euclid-like images have higher S/N than the KiDS-like images,

the AUC results are much better for the latter. After training a model only with r-band images from the KiDS-like data, we can confirm that providing to the network further information, in the form of additional bands, improves its performance. Therefore, the reason why the CNN performs more poorly with these space-based data is that they own merely one band.

Finally, we trained and tested the CNN using more realistic simulated data, that mimic the four bands from Euclid (VIS, J, H, and Y). On this occasion we aimed to accomplish a model with a high precision, since the evaluation method used penalized the amount of false positives. Several models were considered, which were trained with more aggresive definitions for the lenses than that used to evaluate the results. The best model supplied us with a $F_\beta = 0.9859$ (see Fig. 6.35) and an AUC $= 0.700$ (see Fig. 6.34). We succeed in obtaining a pure enough classification of lenses in the challenge, but at the cost of losing more lenses than were detected (see Tabla 6.24). After checking by eye some high scoring false positives, we found that $\approx$33% are true lenses recognizable by eye, but do not meet the evaluation requirements of the challenge. These are encouraging news, since our aggressive models are able to recognize lens patterns in lens examples less clear than those used in the training.

We intend to further explore CNN algorithms, moving on also to the analysis of the properties of the strong lensing events detected. Our long-term goal is to develop automated lens finder models as accurate as possible which are able to extract features difficult to recognize by the human eye. Another important point is ensuring that these models do not depend on the instrument and survey characteristics used to obtain the images. We have seen here that it is possible to apply models trained with KiDS-like data to Euclid-like, SDSS and DES images achieving good results. The models derived need to be robust against small changes in the data sets. We will do this by exploring all the possibilities of CNNs and combining the already detected lenses from different surveys with mock images of strong lenses generated under different conditions.

Chapter 7

Using convolutional neural networks to identify lensed quasars in J-PAS

The morphology of galaxies is a true reflection of the properties of the stars they host (Blanton and Moustakas, 2009; Pozzetti et al., 2010). In the local Universe, elliptical galaxies show larger masses and higher velocity dispersions than spiral galaxies. While the existing stars inside elliptical galaxies tend to be older, giving off more red light than younger stars, spiral galaxies are, due to its higher gas and dust content, hotbeds of star formation. In consequence, the emission from spiral galaxies is brighter and bluer. For this reason, elliptical galaxies can be described as red passive galaxies, and spiral galaxies as blue and active. As more lenses are discovered in new surveys, we are able to better constrain the populations of lenses and sources. The knowledge gathered so far indicates that elliptical galaxies do completely dominate the optical depth for strong lensing by individual galaxies (Turner et al., 1984; Kochanek, 1993).

Normal galaxies are those that contain stars that are generally similar to the stars in our own galaxy, and whose radiation emission is mainly due to their stellar content, dust, and interstellar gas. However, there are galaxies that show additional radiation emission in their SED, not produced by the usual galaxy components. It is thought that in all these galaxies, an enormous amount of energy is released in a tiny nuclear region, so that the origin of this excess radiation is attributed to the existence of an active galactic nucleus (AGN), which is why they are called active galaxies. Therefore, an AGN is a compact region at the center of a galaxy that has a much higher than usual luminosity over some portion of the electromagnetic spectrum. This surplus non-stellar emission has been observed in gamma rays, X-rays, ultra-violet, optical, infrared, microwave and radio. A galaxy with an active nucleus can be considered as the sum of a normal galaxy plus an AGN.

The machinery that powers an AGN is the presence of a supermassive central black hole (SMBH), accreting matter earned from the host galaxy, and converting gravitational energy into electromagnetic radiation. This material forms an accretion disk around the black hole, which is heated by viscosity to temperatures of $\sim 10^6$ K, emitting highly energetic radiation, mainly X-ray and ultraviolet. In addition, internal jets can sometimes emerge perpendicular to the accretion disk, extending beyond the limits of the galaxy itself and ending in radio lobes. This central machinery is surrounded by relatively dense gas clouds that move at very high speeds (broad line region, or BLR). These clouds, in turn, are surrounded by a toroidal structure of gas and dust, much larger than the accretion disk, at moderate temperatures (from $\sim$ 100 K to $\sim$ 1500 K). This toroid is indirectly heated by the AGN radiation, giving rise to a significant emission in the infrared. In turn, it is also surrounded by low-density gas clouds that move at slower speeds than the internal clouds (narrow line region, or NLR).

The observed characteristics of an AGN depend on several properties such as the orientation of the accretion disk regarding our line of sight, the mass of the central black hole or the rate of gas accretion. Based on their observed characteristics, several subclasses of AGNs have been identified, being the quasars one of the most powerful ones. The discovery of quasars (Schmidt, 1963) revealed a class of sources ideal for studying the effects of gravitational lensing. Quasars (from QSOs $\equiv$ "Quasi-Stellar Objects") are distant, and so the probability that they are lensed by foreground galaxies is sufficiently large. Yet, they are intrinsically bright enough to be detected even at cosmological distances. The most distant quasar currently known is at $z = 7.642$ (Wang et al., 2021). Their optical emission region is very compact, much smaller than the typical scales of galaxy lenses. The maximum magnification is inversely proportional to the square root of the size of the object being lensed, so smaller objects can be magnified more than larger objects. Therefore, the resulting magnifications when quasars are lensed can be very large, with well separated and easily detected multiple imaging. Proof of this is the fact that the first lens system ever detected consisted of a two-image lensed quasar Walsh et al. (1979).

Nearly a million of quasars have been found to date, most detected by the Sloan Digital Sky Survey (Lyke et al., 2020). And its amount is expected to increase with the observations to come from the current and future large-scale surveys, which will confirm high-confidence candidates and will bring new detections. A significant number of them will be lensed, and so it is necessary to identify them in an automatic way, that does not require a one-by-one inspection. For instance, the Rubin Observatory/LSST is expected to observe $\sim 10^4$ strongly lensed quasars (Oguri and Marshall, 2010; Verma et al., 2019), or the Euclid survey has a forecast of detecting $\sim 10^3$ lensed quasars (Refregier et al., 2010). J-PAS will observe over half a million quasars, among which we expect hundreds

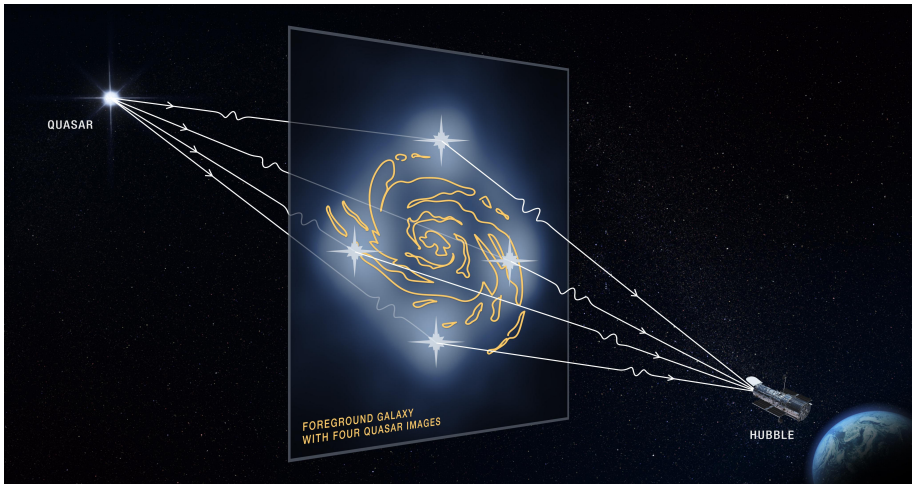


Figure 7.1: Illustration of the light from a distant quasar being disturbed by a massive foreground galaxy. This can result in quadruple images when the background quasar and the foreground galaxy are almost perfectly aligned. Courtesy NASA/JPL-Caltech.

of them to be strongly lensed. Aiming to detect these lensed QSOs in future J-PAS data (Bonoli et al., 2020), we train in this chapter the CNN described in Sec. 5.3 on J-PAS based simulated data.

7.1 The J-PAS survey

The *Javalambre-Physics of the Accelerating Universe Astrophysical Survey* (J-PAS) is an ongoing survey that will cover at least 8000 square degrees of the northern hemisphere extragalactic sky in approximately 5 years, using an innovative system of 54 optical narrow band filters. The data set resulting from this survey will have a unique legacy value, allowing a broad collection of astrophysical studies. The infrastructure required to carry out this large sky photometric survey is placed at the *Observatorio Astrofísico de Javalambre* (OAJ) (see Cenarro et al., 2014). The observatory consists of two telescopes, the Javalambre Survey Telescope (Cenarro et al., 2018, JST/T250) and the Javalambre Auxiliary Survey Telescope (JAST/T80), both located at the Pico del Buitre of the Sierra de Javalambre, Teruel, Spain. Since 2014 the OAJ is a Spanish Singular Scientific and Technical Infrastructure (known with the Spanish acronym ICTS for Infraestructura Científica y Técnica Singular). On the other hand, the exploitation and management of the data produced at the OAJ are liability of the Centro de Estudios de Física del Cosmos de Aragón.

J-PAS will be conducted from the JST/T250, which is an unprecedented Ritchey-Chrétien-like, alt-azimuthal, large-étendue telescope, with 2.55 meters of aperture, and a 3 deg diameter field-of-view (FoV). Its effective collecting area is 3.75 m^2, yielding an étendue of 26.5 m^2 deg^2. However, before the arrival of its main scientific instrument, the Javalambre Panoramic Camera (JPCam), the JST has been equipped with an interim camera, the JPAS-*Pathfinder*, used to carry out the first scientific operations and test the telescope and narrow filter system performances. This camera is composed of a 9.2k×9.2k CCD, with a 0.3 deg^2 FoV and same pixel size as JPCam, 0.2267 arcsec $pixel^{-1}$. The first test accomplished with the JPAS-*Pathfinder* consisted of probing the AEGIS field (along the Extended Groth Strip). The mini-JPAS survey observed 1 deg^2 on the AEGIS field with the 54 J-PAS narrow bands + 2 medium bands + *ugri* bands. All the details concerning the miniJPAS observations and the data set can be found in Bonoli et al. (2020), while the miniJPAS data itself and associated value added catalogs are publicly accessible[1].

The truly novel and distinctive feature of J-PAS lies on its photometric filter system: 54 narrow band filters ranging from 3780 Å to 9100 Å, and two additional broader filters extending to the UV and the near-infrared. The main characteristics of the J-PAS filter set are listed in Table 7.1. The narrow band filters are spaced by about 100 Å, except for the J0378 filter, providing a continuous spectral coverage through the entire optical range. The two additional filters are one medium band covering the UV border (uJAVA), and a broad filter beyond 9100 Å, approaching the near-infrared (J1007). This large number of filters will allow resolving spectral features such as the Lyman-α line of high redshift quasars, and other strong emission and absorption lines. Fig. 7.2 shows the transmission curves of these 56 filters, after taking into account the atmospheric absorption, the CCD efficiency and the camera optics. The miniJPAS field has also been observed with SDSS-like broad band filters: u, g, r, and i. The average FWHM values of the point spread functions (PSFs) per tile and per filter are shown in Fig. 7.3.

The J-PAS filter system definition was determined to pursue three main scientific goals. First, to measure photometric redshifts for galaxies up to $z \sim 1$ with a high precision of $\sim 0.3\%$. Second, to study stellar populations in nearby galaxies. And finally, to resolve broad spectral features of objects such as supernovae and AGNs. The size of the JPAS-*Pathfinder* camera FoV allowed to cover the AEGIS field almost entirely with only 4 pointings, being the observations carried out with the instrument rotated $\sim 45°$ with respect to the celestial North. The four tiles composing the miniJPAS footprint are listed in Table 7.2 and are shown in Fig. 7.4, along with the observational fields of other overlapping survey projects.

[1] http://archive.cefca.es/catalogues/minijpas-pdr201912

Filter	λ (Å)	$\delta\lambda$ (Å)	Filter	λ (Å)	$\delta\lambda$ (Å)
uJAVA	3497	509	J0650	6506	146
J0378	3782	157	J0660	6607	151
J0390	3904	150	J0670	6710	146
J0400	3996	148	J0680	6812	152
J0410	4110	144	J0690	6912	148
J0420	4203	147	J0700	7007	148
J0430	4303	148	J0710	7119	148
J0440	4403	152	J0720	7207	146
J0450	4503	150	J0730	7307	150
J0460	4603	148	J0740	7414	148
J0470	4701	148	J0750	7502	142
J0480	4799	142	J0760	7602	144
J0490	4902	154	J0770	7719	146
J0500	5002	152	J0780	7811	145
J0510	5097	148	J0790	7907	142
J0520	5202	150	J0800	8009	140
J0530	5296	150	J0810	8124	144
J0540	5389	152	J0820	8226	143
J0550	5498	149	J0830	8329	148
J0560	5596	150	J0840	8429	148
J0570	5701	150	J0850	8523	146
J0580	5803	148	J0860	8620	148
J0590	5917	152	J0870	8716	146
J0600	6010	150	J0880	8810	146
J0610	6107	150	J0890	8912	150
J0620	6206	148	J0900	9000	154
J0630	6309	150	J0910	9107	152
J0640	6408	150	J1007	9316	635

Table 7.1: J-PAS photometric filter system main characteristics.

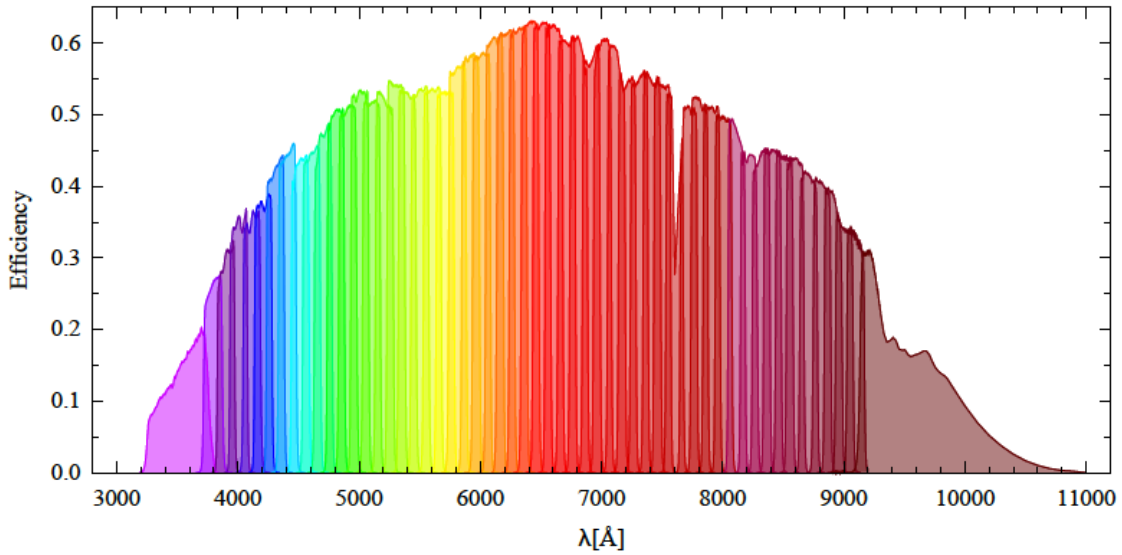


Figure 7.2: Transmission curves of the J-PAS filters. Taken from Bonoli et al. (2020).

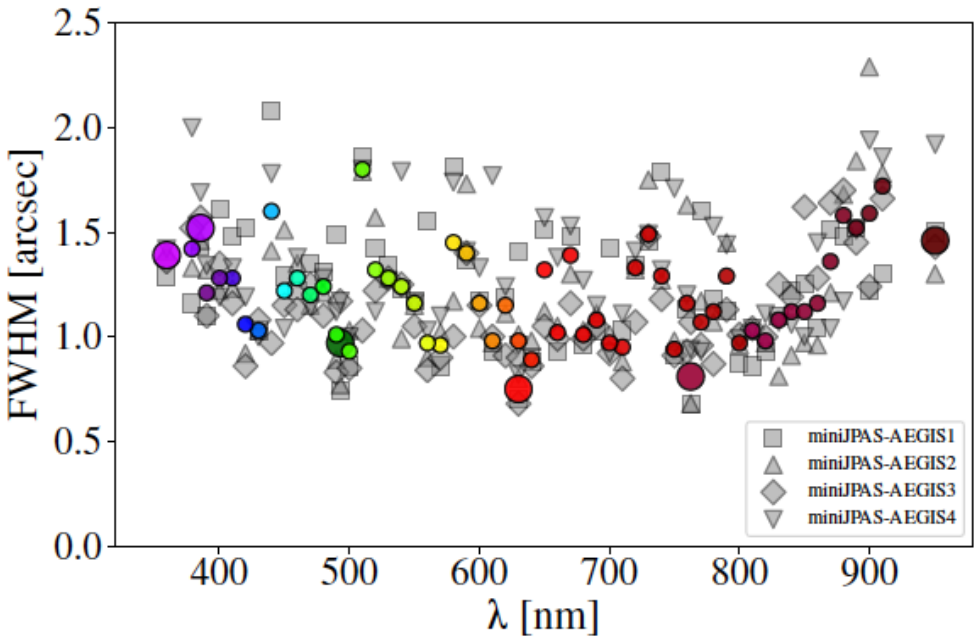


Figure 7.3: PSF FWHM values per tile and per filter. The colored symbols represent the average values for each filter, while the gray ones are the values for each one of the pointings. The larger symbols indicate the FWHM of the broad bands (uJAVA, u_{JPAS}, g_{SDSS}, r_{SDSS}, i_{SDSS}, J1007). Taken from Bonoli et al. (2020).

Tile	RA J2000 (deg)	Dec J2000 (deg)
miniJPAS-AEGIS1	214.2825	52.5143
miniJPAS-AEGIS2	214.8285	52.8487
miniJPAS-AEGIS3	215.3879	53.1832
miniJPAS-AEGIS4	213.7417	52.1770

Table 7.2: Central coordinates of each of the tiles composing the miniJPAS footprint.

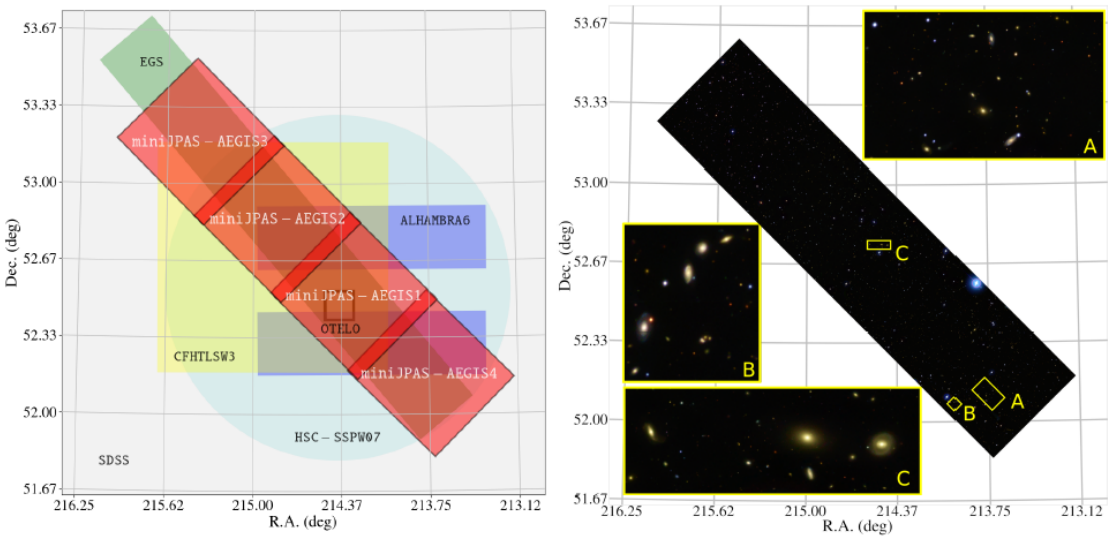


Figure 7.4: *Left*: Footprint of the miniJPAS survey, showing also its overlap with other surveys. The tiles that compose it are depicted as red squares. Their central coordinates can be seen in Table 7.2. *Right*: *g-r-i* color image of the miniJPAS footprint. zooming over three selected areas. Taken from Bonoli et al. (2020).

7.2 Mini-JPAS data

In this chapter, we aim for simulating strongly lensed quasars in J-PAS, so as to be used for training the CNN architecture explained in Sec. 5.3. Whereas the lensed quasars are based on a mock catalog, real red galaxies observed with J-PAS are assumed as lenses. The reason for using red galaxies is that, due to their mostly higher surface mass density, they lead more often to lensing events. The characteristics of the data (real and mock) used to train the network are told in this section, and the process for creating the lensing simulations is described in Sec. 7.3.

7.2.1 Galaxies in the mini-JPAS footprint

We consider a list of 4139 galaxies, with redshifts between 0.15 and 0.5, found in the mini-JPAS footprint (Bonoli et al., 2020). Their corrected spectra were fitted by the SED fitting codes BaySeAGal (de Amorim et al. in prep) and MUFFIT (Díaz-García et al., 2015). BaySEAGal is a Bayesian inference code for spectral fitting which uses the data from the 54 narrow-band and 2 wide-band filters in J-PAS, and assumes a τ-delayed star formation history (SFH). It allows deriving a wide variety of intrinsic galaxy properties, such as the stellar mass, the luminosity-weighted age of the stellar population, the metallicity and stellar extinction. On the other hand, MUlti-Filter FITting (MUFFIT) is a generic spectral fitting code, based on a error-weighted χ^2, optimized to retrieve the main stellar population parameters of galaxies in photometric multi-filter surveys by assuming two single stellar populations models. Our estimation of the stellar mass for these galaxies is the average of the corresponding derivations from both codes. We assume as red galaxies those identified as such by both codes, attending to their color-magnitude properties, while the rest will be considered blue galaxies. This yardstick leaves us with 650 red galaxies and 3489 blue galaxies. We downloaded squared images (16" $\times$ 16") of these galaxies centered at the positions provided by the catalog. Only the red galaxies are used to perform realistic simulations of the gravitational lenses. However, not all red galaxies selected have photometric flux measurements in all bands, which is necessary to carry out the lensing simulations. For this reason, we divide the sample of elliptical galaxies that does have this information, allocating half (207) for the lensing simulations and the other half (206) to be non-lensing examples. Red galaxies that lack of photometric information in any of the bands (237) and non-elliptical galaxies are purposed to be non-lensing examples. Besides, the size of the sample of non-lensing galaxies is doubled by rotating every galaxy 90°, resulting in 7864 examples.

7.2.2 Mock J-PAS based quasars

We use a mock quasars catalog with fluxes simulated based on the J-PAS photometric system. These mocks were created by convolving real SDSS spectra, selected from the DR12 quasar catalog (Pâris et al., 2017), with the mini-JPAS photometric passbands, and adding J-PAS-like noise, in order to achieve the expected S/N for J-PAS. We considered realistic photometric errors (within an aperture of 3") selected from a catalog of point-like sources observed with mini-JPAS. The quasar distribution was derived from theoretical predictions based on the luminosity function (Palanque-Delabrouille et al., 2016) and corresponds to

an area of approximately 300 $\deg^2$. The catalog contains 105,846 quasars. In order to have a sample of quasars well distributed in magnitudes, four versions of the quasar catalog were created. Each individual source appears in all versions, but its magnitude is different. Version 1 (v1) corresponds to the original r-band magnitudes, after shifting the original SDSS magnitude distribution to fainter values, so that it resembles the distribution expected for J-PAS. Versions v2, v3 and v5 were derived from v1 by summing, respectively, one, two and four magnitudes to the r-band values in the reference v1 version.

7.2.3 Quasars and stars in the mini-JPAS footprint

As non-lensing examples, along with the already discussed sample of galaxies, we have also trained the network with quasars and stars, identified in the mini-JPAS footprint, in order to get more robust results. Including non-lensed quasars in the training sample is important since most of the quasars observed with J-PAS will not be lensed. It is also worth to mention that blue stars are easily mistaken for quasars. We consider a list of 131 quasars and 111 stars with spectroscopic confirmation in the SDSS DR16 Superset catalog (Lyke et al., 2020). Note that this is a very reduced list due to two reasons. The Superset catalog contains a small set of sources targeted as quasars for eBOSS, with great completeness and purity due to visual inspection procedures to correct the automated classifications and redshifts from the pipeline. In addition, our sample was obtained by doing a cross-match within 1" with the Superset catalog, in which only miniJPAS sources with FLAGS $= 0$, $\mathrm{MASK}_{\mathrm{FLAGS}} = 0$ and $z_{\mathrm{QUALITY}} = 3$ were considered. Although there are fewer objects than we would obtain by doing a cross-match with the SDSS pipeline, the classification of these objects is more reliable. 36 of the quasars in our sample were obtained after searching specifically for quasars with $r > 22$.

7.3 Lensed quasars simulations

Once all data have been detailed, we can proceed to explain the lensing simulations we carried out. Through this work, we adopt a flat cosmological model with $\Omega_M = 0.3$, $\Lambda = 0.7$ and $h = 70\ \mathrm{Km\,s^{-1}Mpc}$. We assume that the elliptical galaxies we will use as lenses behave as softened isothermal ellipsoids. Following this assumption, the mass distribution of the galaxies is modeled using Eq. (3.41), and making the variable change shown in Eq. (3.47) to account for the ellipticity. The stellar mass, redshift, ellipticity, and orientation angle of each galaxy are directly read from the catalog. In order to augment the number of

examples, six different lenses are built for every individual galaxy used. This is done by considering 3 different masses: its estimated value from photometric measurements (m), double its value ($2m$), and half its value ($\frac{1}{2}m$). And 2 rotation angles: its orientation angle and rotating the galaxy 90°. Since 207 red galaxies are being used to create the lensing simulations, we can build up to 1242 (207 × 2 rotations × 3 masses) different lens models. The deflection angle for these lens models is calculated according to Eq. (3.11).

Quasars are built occupying one single pixel (1 pixel = 0.01 arcsec) in the source plane at the redshift read from the mock quasar catalog, and they are placed randomly in positions close to caustics, so that the magnification of the source is above a certain threshold. From the full mock quasar catalog we are using in the lensing simulations those that fall in the redshift range $0.8 < z < 2.5$ (76,626 quasars), to roughly overlap with the range where the number of expected quasars is highest (Pâris et al., 2018). The total magnification of the quasar is estimated as the ratio between the area of the lensed images in the lens plane and the area of the unlensed quasar, which fills a single pixel, in the source plane. Once the quasar is lensed, we estimate its new flux in the r band by multiplying its tabulated flux in this band by the total magnification. As we are using the tabulated fluxes from the mock quasar catalog, and we do not want the noise to be magnified along with the signal, instead of multiplying their fluxes at all bands by the total magnification, we replace the lensed quasar with another mock quasar close enough both in redshift and in the r-band magnitude. We can do this easily thanks to having created several versions of the quasar catalog, which contain the same quasars but simulated at different magnitudes.

Not all simulations performed are finally accepted and used for training the network. If no surrogate quasar is found that meet the conditions to replace the lensed one, the simulation is rejected. In addition, we demanded the simulations to meet at least one of the two following requirements in order to be accepted. The lensed quasar and the lens galaxy have to be either close in the r-band magnitude ($r_{\text{lensed qso}} - r_{\text{elliptical galaxy}} < 2.6$), or the lensed quasar must have a threshold magnification of $\mu_{\text{total}} > 5$. The simulations accepted move on to the following step. The images in all J-PAS filters of the corresponding galaxy, whose data have been used to model the lens, are either rotated or left as they are so as to match the orientation of the lens in the simulation. Using observed galaxies allows us to include real systematics, such as instrumental noise and the PSF of the instrument, in the simulations. We normalize the total flux of these images so they match the tabulated fluxes. On the other hand, we convolve the map of the lensed quasar with the corresponding PSF in each J-PAS channel (see Fig. 7.3), and repixelize the resulting maps so they have the same pixel size as the J-PAS images. We then normalize the overall fluxes within these images so they correspond to the fluxes in the catalog. Finally, we add up at each channel the map of the lensed quasar with the image of the elliptical galaxy. As it is shown in

Fig. 7.5, every lensed quasar resulted from this process, and the non-lensing examples too, can be seen as a cube, whose height and width correspond to the side sizes of the cutout images, and its depth is equal to the number of J-PAS filters considered. Using this cube as input for the convolutional neural network implies working with a large volume of data. This would result in a more complex network, i.e., with a higher number of parameters, and a slower learning process. In order to avoid this issue, we reduce this cube to only two images (hereafter nicknamed pseudo-spectra, since they mimic observations done with a narrow slit, with the spectral resolution of the J-PAS filter system) by projecting the data along the height and width axes for all the depth layers. The dimensions of these resulting pseudo-spectra correspond to the side sizes of the images in the projected axis and the number of channels. These pseudo-spectra contain all the relevant information about the morphological and spectral features of both, the simulated lensed quasars and the non-lensing examples. Some examples of the lensed quasar simulations at specific wavelengths can be seen in Fig. 7.6. The corresponding pseudo-spectra for these examples are shown in Fig. 7.7. On the other hand, some examples of pseudo-spectra for non-lensed quasars and stars are shown in Fig. 7.8, and examples for blue and red galaxies can be seen in Fig. 7.9.

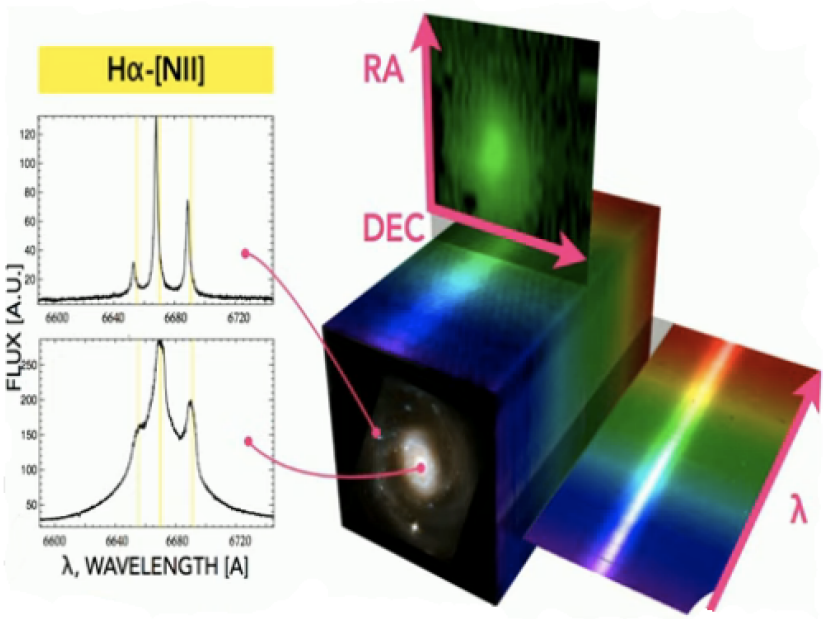
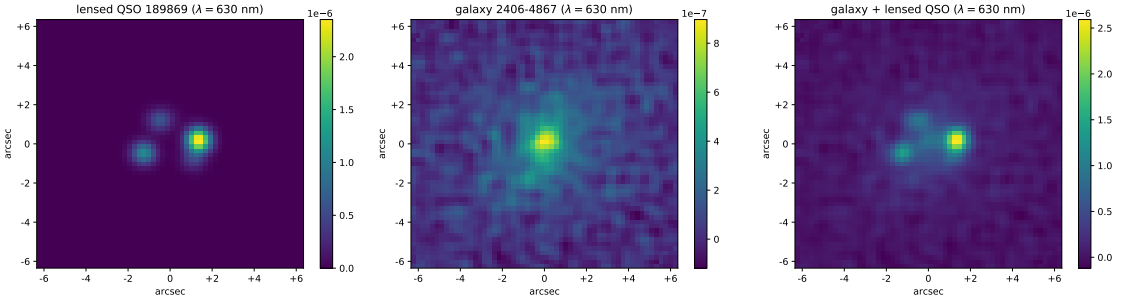
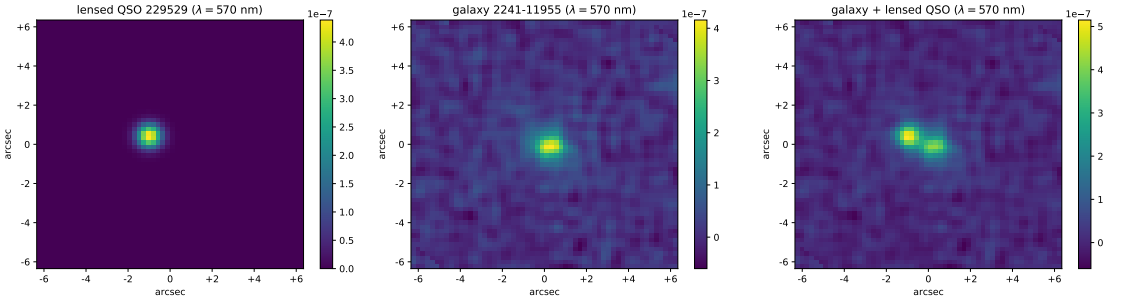


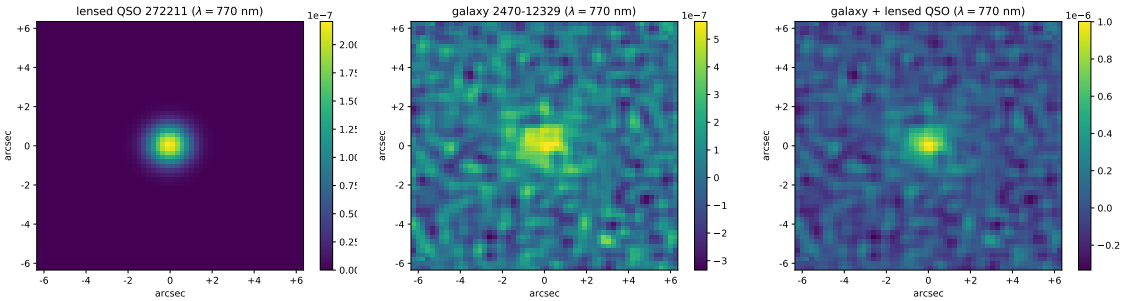
Figure 7.5: llustration showing the construction of the pseudo-spectra. Taken from a talk given by a researcher of the Instituto de Astrofísica de Andalucía.



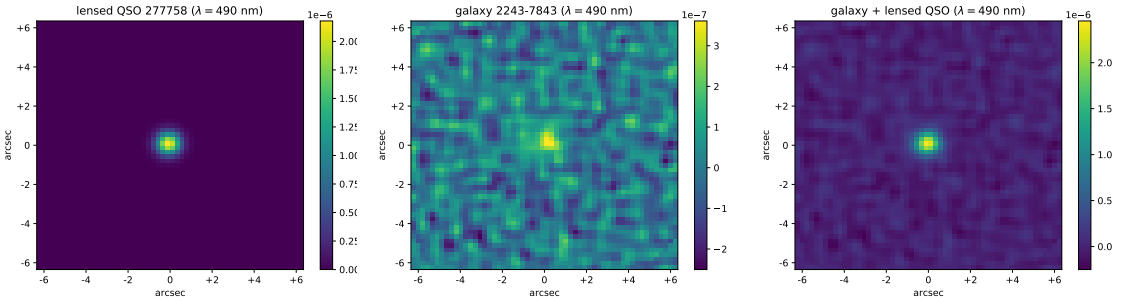
(a) QSO (189869, $z = 2.16$) (b) Lens (2406-4867, $z = 0.19$) (c) QSO + lens at $\lambda = 630$ nm



(d) QSO (229529, $z = 1.43$) (e) Lens (2241-11955, $z = 0.26$) (f) QSO + lens at $\lambda = 570$ nm

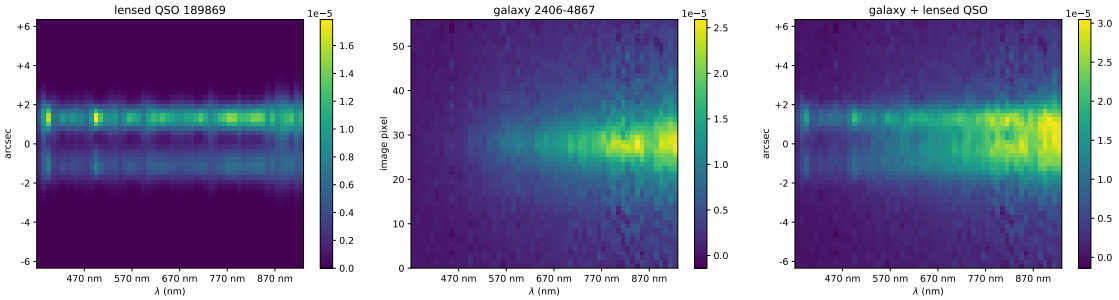


(g) QSO (272211, $z = 0.94$) (h) Lens (2470-12329, $z = 0.30$) (i) QSO + lens at $\lambda = 770$ nm

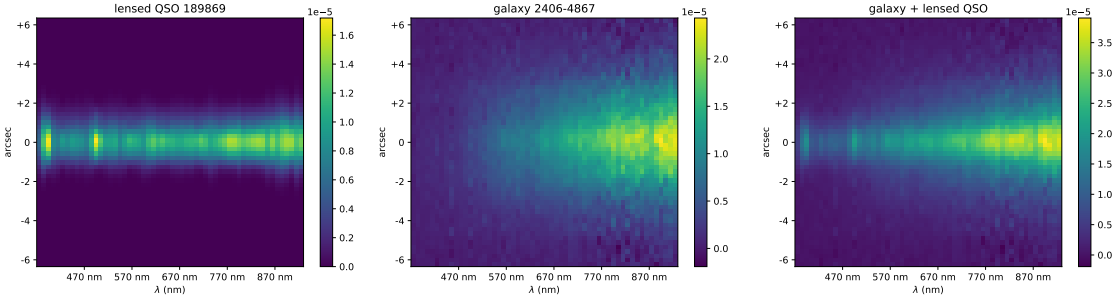


(j) QSO (277758, $z = 1.78$) (k) Lens (2243-7843, $z = 0.27$) (l) QSO + lens at $\lambda = 490$ nm

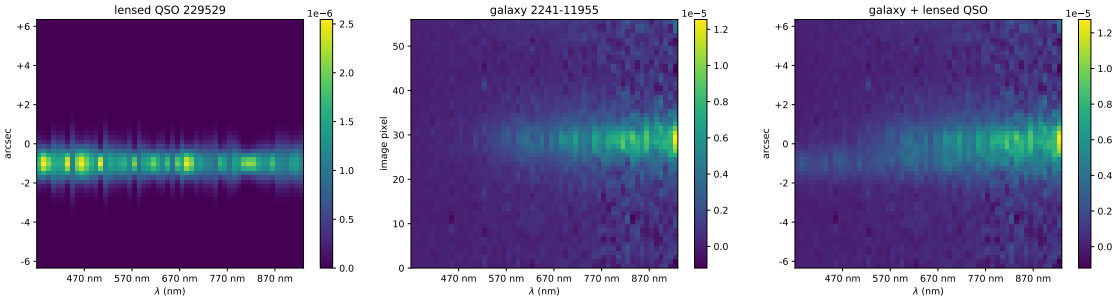
Figure 7.6: Some examples of the lensed QSOs simulations. Every row represents a different simulation at a specific wavelength (λ), showing the lensed QSO (*left*), the lens galaxy (*middle*), and the QSO + lens resulting image (*right*).



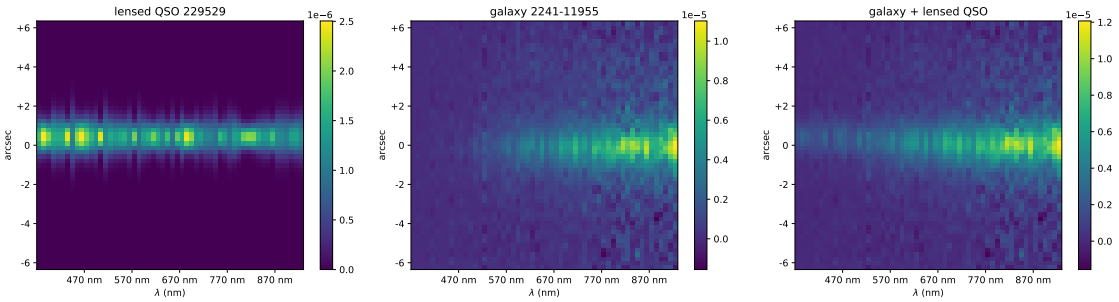
(a) QSO 189869 (x-axis) (c) Lens 2406-4867 (x-axis) (e) QSO + lens (x-axis)



(b) QSO 189869 (y-axis) (d) Lens 2406-4867 (y-axis) (f) QSO + lens (y-axis)

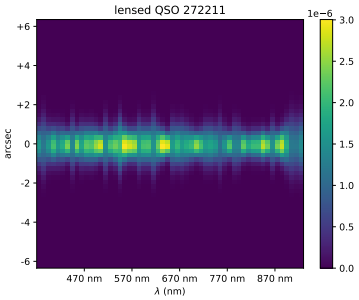


(g) QSO 229529 (x-axis) (i) Lens 2241-11955 (x-axis) (k) QSO + lens (x-axis)

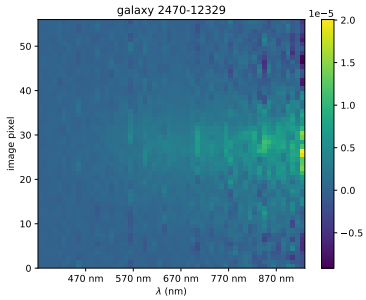


(h) QSO 229529 (y-axis) (j) Lens 2241-11955 (y-axis) (l) QSO + lens (y-axis)

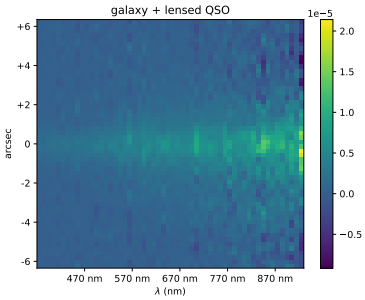
Figure 7.7: Some pseudo-spectra examples of the lensed QSOs simulations. Every simulation spans two rows, with the x-axis/y-axis pseudo-spectra for the lensed QSO (*left*), the lens galaxy (*middle*), and the QSO + lens image (*right*) shown in the upper/lower row.



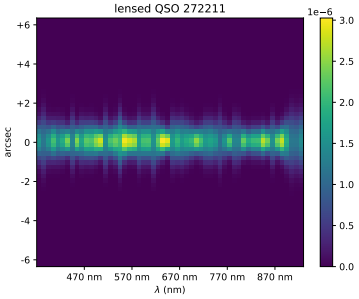
(m) QSO 272211 (x-axis)



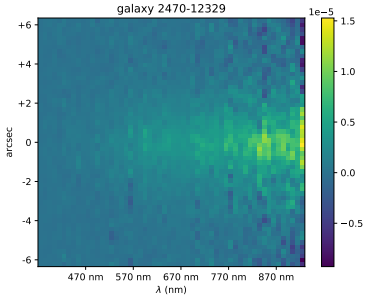
(o) Lens 2470-12329 (x-axis)



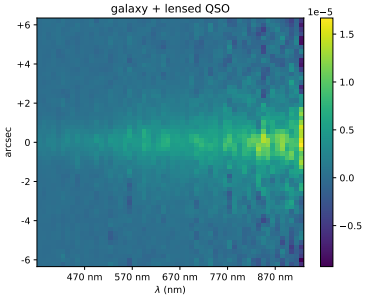
(q) QSO + lens (x-axis)



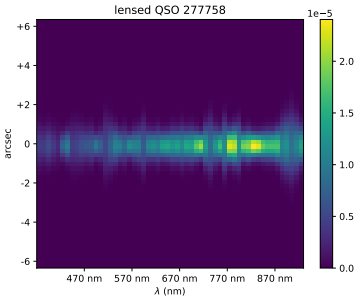
(n) QSO 272211 (y-axis)



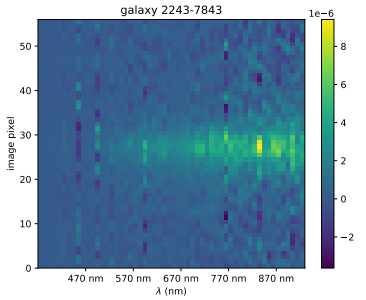
(p) Lens 2470-12329 (y-axis)



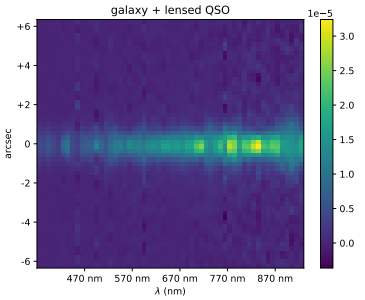
(r) QSO + lens (y-axis)



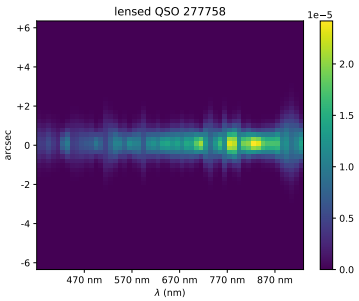
(s) QSO 277758 (x-axis)



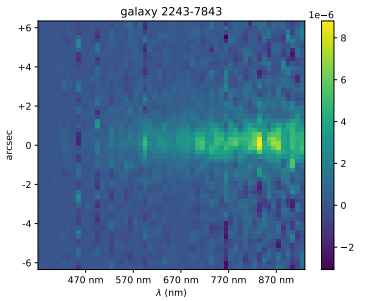
(u) Lens 2243-7843 (x-axis)



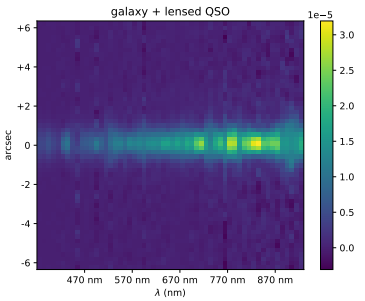
(w) QSO + lens (x-axis)



(t) QSO 277758 (y-axis)

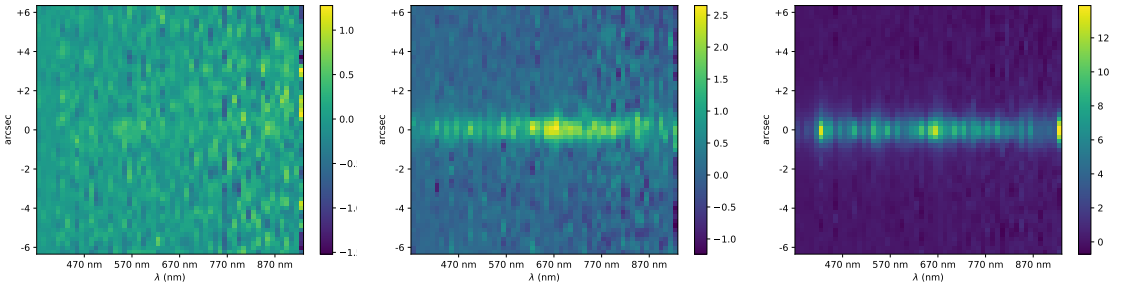


(v) Lens 2243-7843 (y-axis)

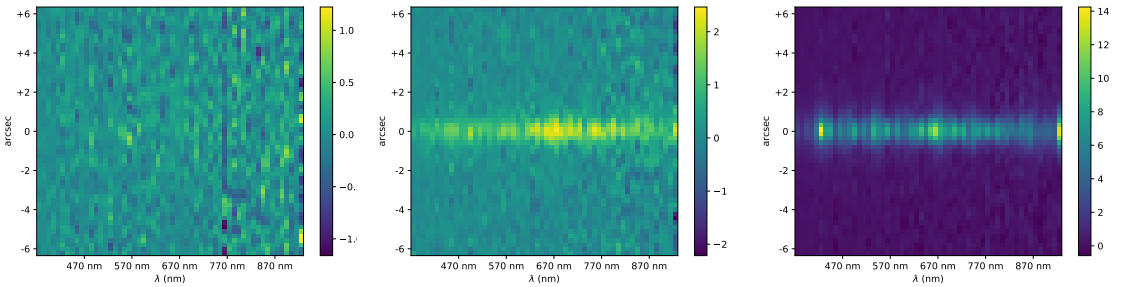


(x) QSO + lens (y-axis)

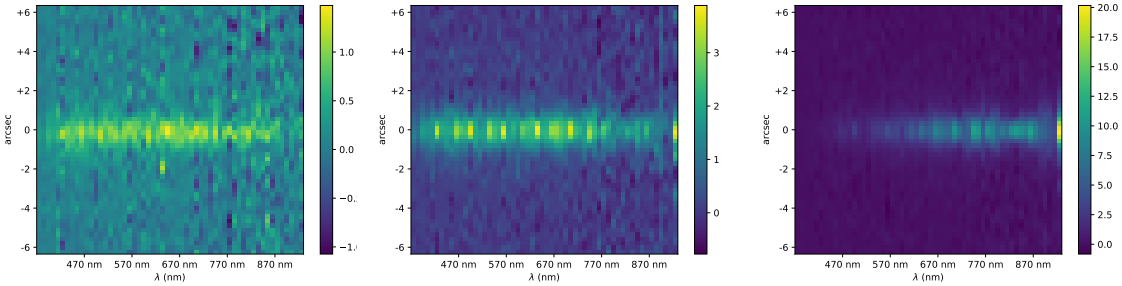
Figure 7.7 (cont.)



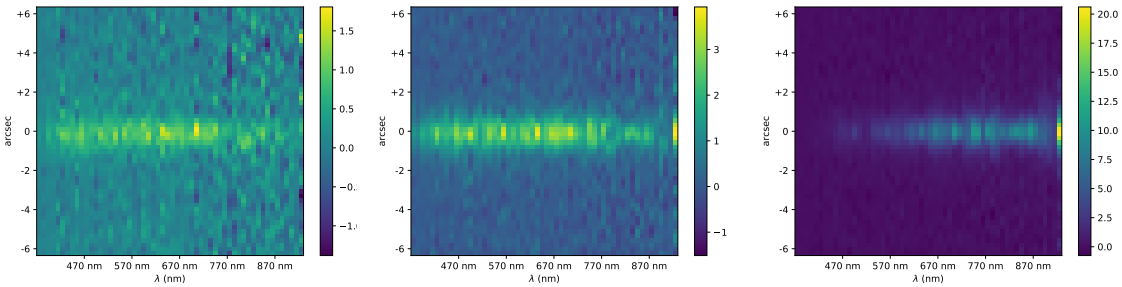
(a) QSO 2243-9040 (x-axis) (c) QSO 2406-5992 (x-axis) (e) QSO 2406-11608 (x-axis)



(b) QSO 2243-9040 (y-axis) (d) QSO 2406-5992 (y-axis) (f) QSO 2406-11608 (y-axis)

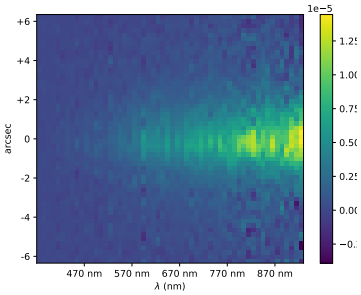


(g) star 2470-13619 (x-axis) (i) star 2243-2625 (x-axis) (k) star 2243-14817 (x-axis)

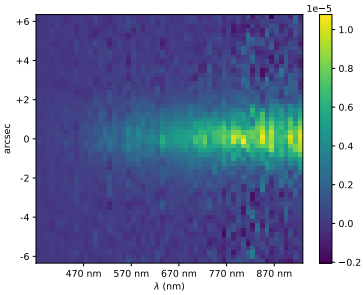


(h) star 2470-13619 (y-axis) (j) star 2243-2625 (y-axis) (l) star 2243-14817 (y-axis)

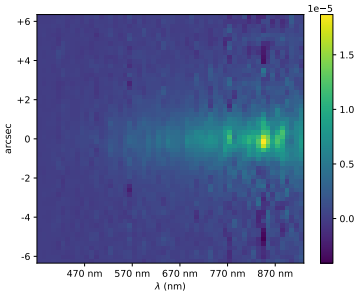
Figure 7.8: Some pseudo-spectra examples of non-lensed QSOs and stars. Every example spans two vertical boxes, with the x-axis/y-axis pseudo-spectrum above/below.



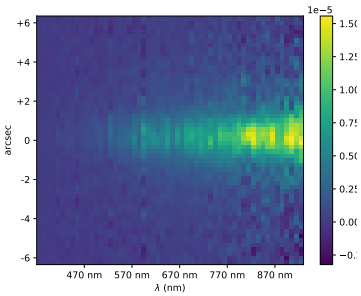
(a) 2241-7123 (red, x-axis)



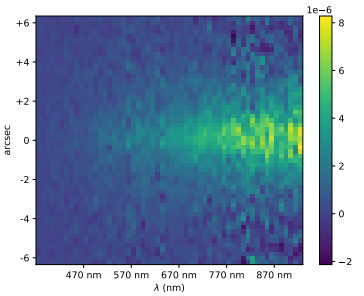
(c) 2406-7091 (red, x-axis)



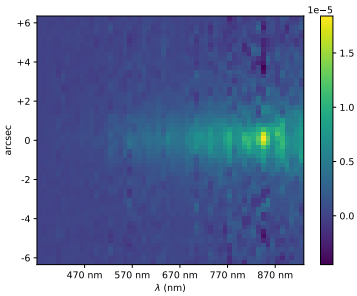
(e) 2470-1055 (red, x-axis)



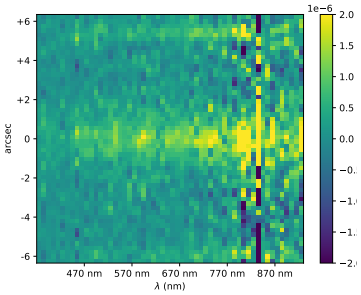
(b) 2241-7123 (red, y-axis)



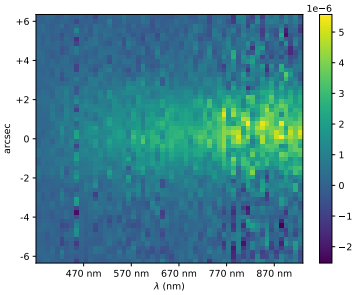
(d) 2406-7091 (red, y-axis)



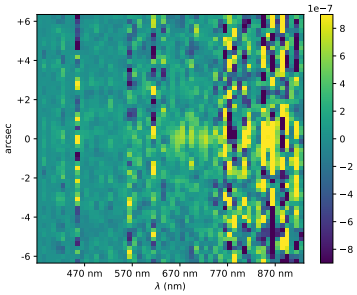
(f) 2470-1055 (red, y-axis)



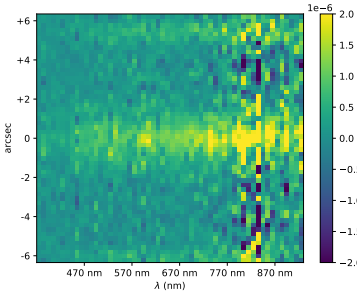
(g) 2406-4876 (blue, x-axis)



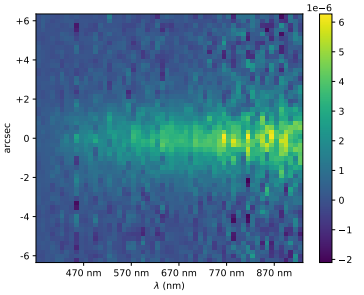
(i) 2406-12109 (blue, x-axis)



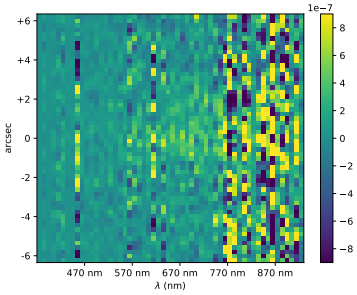
(k) 2470-12036 (blue, x-axis)



(h) 2406-4876 (blue, y-axis)



(j) 2406-12109 (blue, y-axis)



(l) 2470-12036 (blue, y-axis)

Figure 7.9: Some pseudo-spectra examples of red and blue galaxies. Every example spans two vertical boxes, with the x-axis pseudo-spectrum above and the y-axis pseudo-spectrum below. Vertical bands are due to noisier images for those specific filters.

7.4 Results

We have trained the network using two data sets, whose only difference is that with the first one we are not showing to the network examples of unlensed QSOs and stars. 10% of the examples in both training sets were used for validation. Data from every example were normalized to the maximum value in each pseudo-spectrum before being feed up to the network. The network architecture used is the same as described in Sec. 5.3, but again with the exception that input data consist of two channels, namely the pseudo-spectra in the x and y projections. We use early-stopping, so the learnings with the training data sets concluded after ~ 50 epochs. The number of trainable parameters of the network per each layer for the trainings carried out are shown in Table 7.3. The models derived from these trainings were applied to the same test data set, which includes unlensed QSOs and stars. The characteristics of the data sets considered here are summarized in Table 7.4.

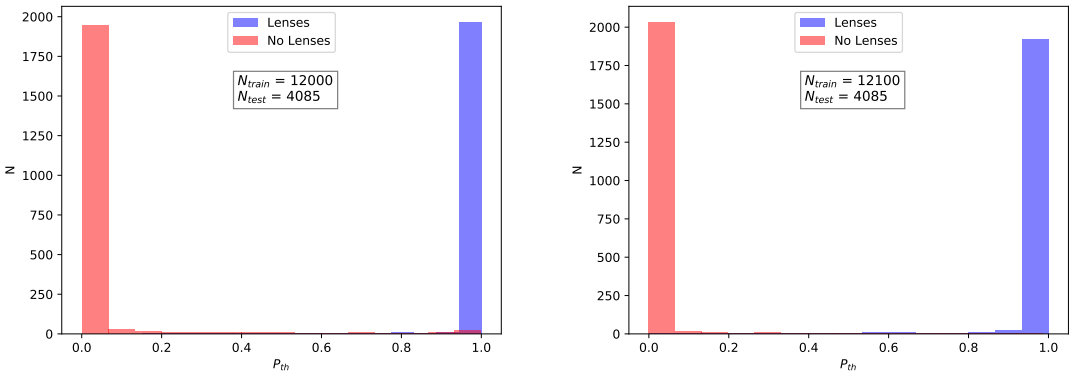
As we have said, the models differ just in the fact that one of them has been trained with QSOs and stars, as non-lensing examples, while the other one has not. The probability distributions and confusion matrices of these models on the test data set are shown in Fig. 7.10 and Table 7.5, respectively. The separation achieved between classes is astonishing good, so it would not be necessary to go beyond the default probability threshold of 0.5 to get a pure sample of lensed QSOs. Looking in more detail, training with unlensed QSOs and stars has a significant positive impact, decreasing the amount of false positives from 61 examples (2.9%) to just 9 (0.4%). These 9 examples include 3 blue galaxies, 2 QSOs, and 3 stars. The list of false positives can be read in Table 7.6, where these 9, common to both models, are marked with an asterisk. The minor side effect is that the false negatives increase from 6 (0.3%) to 24 (1.2%). Although the TPR drops, the accuracy and precision grow with this decision. The ROC curves shown in Fig 7.11 confirm the logical conclusion that training with a greater variety of examples yields more reliable results.

Layer	Number of weights
1st convolutional layer	2,336
2nd convolutional layer	51,264
3rd convolutional layer	32,896
4th convolutional layer	147,584
1st fully connected layer	1,605,696
2nd fully connected layer	65
Full network	1,839,841

Table 7.3: Number of weights per layer of the network trained with J-PAS based data.

data set	Size	Composition
Training data set 1	13,359	7460 lensed QSOs 5234 blue galaxies 665 red galaxies
Training data set 2	13,481	7460 lensed QSOs 5234 blue galaxies 665 red galaxies 66 QSOs 56 stars
Test data set	4085	2000 lensed QSOs 1744 blue galaxies 221 red galaxies 65 QSOs 55 stars

Table 7.4: Description of the two J-PAS based training data sets used, and the only test data set considered.



(a) QSOs and stars excluded from training.

(b) QSOs and stars included in training.

Figure 7.10: Probability distributions of the models derived with the CNN using J-PAS based lensing simulations, when applied to the same test sample. Fig. 7.10a shows the results of a model trained considering only galaxies as non-lensing examples, while in Fig. 7.10b the model was trained including also QSOs and stars as non-lensed examples. Blue bins repesent those examples in the test sample that are lenses and the red bins correspond to the non-lenses.

True \ Predicted	Non lens	Lens
Non lens	2024 (97.1%)	61 (2.9%)
Lens	6 (0.3%)	1994 (99.7%)

(a) QSOs and stars not part of training. ($Acc = 98.4\%$. $Pre = 97.0\%$)

True \ Predicted	Non lens	Lens
Non lens	2076 (99.6%)	9 (0.4%)
Lens	24 (1.2%)	1976 (98.8%)

(b) QSOs and stars included in training. ($Acc = 99.2\%$. $Pre = 99.5\%$)

Table 7.5: Confusion matrices of the models derived with the CNN using J-PAS based lensing simulations, when applied to the same test sample. Table 7.5a shows the results of a model trained considering only galaxies as non-lensing examples, while in Table 7.5b the model was trained including also QSOs and stars as non-lensed examples. $P_{thr} = 0.5$ is chosen as the reference probability threshold. The accuracy (Acc) and precision (Pre) of the different models are also calculated.

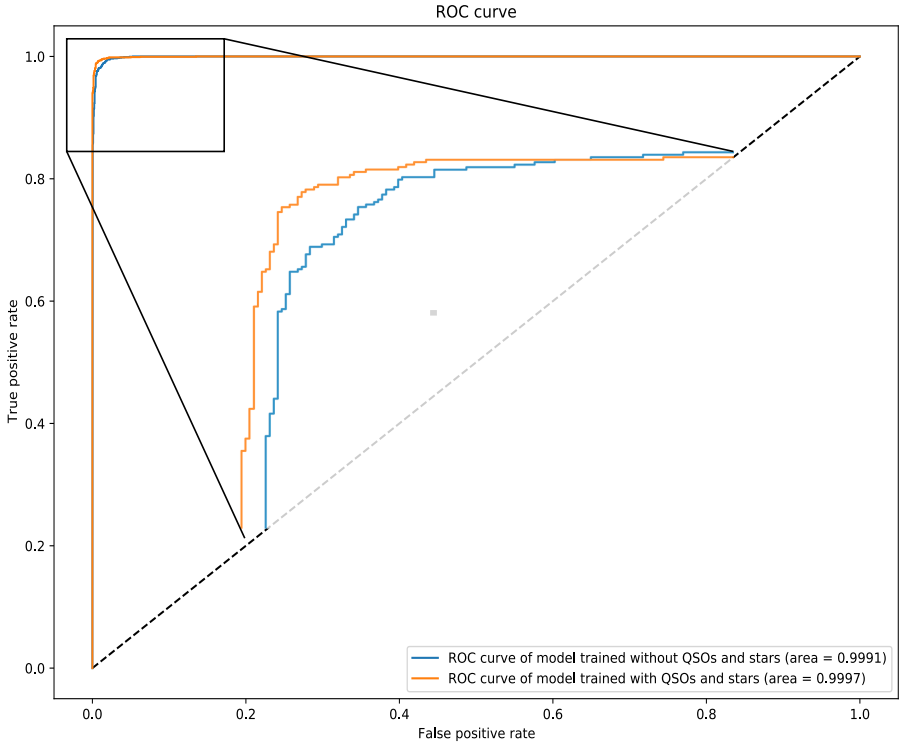


Figure 7.11: ROC curves obtained by applying the models derived with the CNN using J-PAS based lensing simulations to the same test sample. The model trained considering only galaxies as non-lensing examples is depicted with a blue line while the orange line represents the model trained including also QSOs and stars as non-lensed examples.

No.	Type	ID	P_{lens} (training data set 1)	P_{lens} (training data set 2)
1	star	2241-19461	0.500	0.040
2	red gal	2406-3601-2	0.503	0.354
3	blue gal	2241-16211-2	0.508	0.006
4	blue gal	2241-8258-1	0.514	0.003
5	blue gal	2406-2726-1	0.518	0.006
6	blue gal	2241-9424-1	0.522	0.324
7	qso	2406-5548	0.523	0.001
8	blue gal	2470-12555-1	0.574	0.021
9	blue gal	2243-3842-1	0.595	0.024
10	qso	2406-1224	0.609	0.001
11	qso	2241-18457	0.611	0.409
12	star	2243-9710	0.621	0.024
13	blue gal	2406-7789-1	0.622	0.002
14	qso	2241-15867	0.635	0.020
15	qso	2241-19660	0.648	0.001
16	blue gal	2243-1682-2	0.670	0.016
17	blue gal	2470-7406-2	0.682	0.035
18	qso	2406-11219	0.696	0
19*	blue gal	2241-4599-2	0.710	0.765
20	blue gal	2243-8607-1	0.714	0.284
21	star	2241-5588	0.718	0.035
22	qso	2406-9169	0.733	0.004
23	blue gal	2241-6367-2	0.791	0.149
24	blue gal	2243-6289-1	0.809	0.227
25	blue gal	2470-14948-1	0.810	0.071
26	red gal	2406-9636-1	0.818	0.148
27	blue gal	2406-9456-1	0.827	0
28	qso	2243-12523	0.828	0.028
29	qso	2241-1126	0.853	0.486
30	blue gal	2470-1625-2	0.872	0.029
31	blue gal	2241-4599-1	0.878	0.209
32	star	2243-8359	0.885	0
33	blue gal	2406-2674-1	0.901	0.011
34	red gal	2406-10757-2	0.909	0.001
35	qso	2241-13090	0.909	0.020
36	qso	2470-8427	0.923	0.275
37	red gal	2406-2835-1	0.944	0.265
38*	blue gal	2241-9990-1	0.949	0.797
39	red gal	2406-5738-1	0.953	0.418
40*	blue gal	2243-9957-1	0.953	0.890
41	star	2241-6436	0.956	0.073
42	blue gal	2243-6289-2	0.960	0.301
43	blue gal	2406-16273-2	0.960	0.433
44	star	2243-12739	0.960	0.036
45	qso	2243-6718	0.968	0.003
46	star	2406-7843	0.982	0.278
47	star	2243-3813	0.983	0.187
48	blue gal	2241-19534-2	0.984	0.115
49	star	2243-2625	0.988	0.186
50	blue gal	2406-3603-1	0.993	0.002
51	star	2241-19071	0.994	0.040
52	star	2241-10309	0.995	0.258
53	qso	2243-12132	0.999	0.056
54	star	2243-14817	0.999	0.260
55	blue gal	2241-20674-1	1	0.012
56*	qso	2243-12507	1	0.685
57*	star	2470-12146	1	0.640
58*	star	2406-1001	1	0.985
59*	star	2241-8822	1	0.972
60*	qso	2241-3062	1	0.854
61*	star	2243-6036	1	0.982

Table 7.6: False positive examples in the J-PAS based test sample, i.e., non-lenses predicted as lenses by our models. The 9 false positives from the model trained including non-lensed QSOs and stars (training data set 2) were also missclassified by the other model. Those examples missclassified by both models are identified with an asterisk. The probability scores from both models for every example are shown.

7.5 Conclusions

In this chapter the first results of an ongoing work aimed to develop an automatic identification method for lensed quasars in J-PAS are shown. We have created strong lensing simulations using mock quasars based on the J-PAS photometric system, and red galaxies observed in the mini-JPAS footprint. These lensing simulations, along with non-lensing examples (galaxies, quasars and stars), were used to train a CNN. Instead of training the network with the images of the 56 bands considered, we used two pseudo-spectra for every example, which were built by projecting the data in all bands along the x and y axes of the images. We may loss some information, but the relevant morphological and spectral features of the examples needed to classify them are mostly preserved, and the complexity of the network is less. We compared the performance of two models, one trained with galaxies, quasars, and stars, as non-lensing examples, and the other trained just with galaxies. We find that both models achieve impressive results, but training with a greater diversity of examples reduces meaningly the amount of false positives. We are aware of the fact that the goodness of the results is mostly due to the low number of QSOs/stars in the test sample, and the requirements under which the lensing simulations have been built. Nevertheless, our desire is to check next the performance of the CNN by increasing the amount of low-μ_{eff} lensed quasars, which will be surely less easily classifiable. Increasing the numbers of stars, quasars, and galaxies considered here is also in our work path. For the future, we aim to test the models derived on the sources detected in the full J-PAS footprint. The correct identifications of lensed quasars achieved will be of interest for the WEAVE-QSO survey (Pieri et al., 2016), whose targets will be J-PAS quasars.

Chapter 8

Conclusions and future work

It has not been that long since convolutional neural networks started to be used in cosmology, but they have accomplished to become one of the most attractive tools for analyzing data, in particular for clasification purposes. And they will surely continue to be increasingly relevant due to their ability to detect features without any human supervision. We have shown in this work how they can be successfully employed for searching strong lenses in astronomical images. But they are as well used for distinguishing between different astronomical sources or radiations, in morphological classification of galaxies, or for measuring photometric redshifts. For all these purposes, having a good knowledge of the observed SED is crucial so as the network will be able to characterize the nature or estimate the redshift of the detected objects. Although spectroscopic surveys allow accurate analyses of the sources, lengthy follow-up campaigns are required for every target. On the other hand, wide photometric surveys provide data for as many sources as fall in the images, but the wavelength resolution depends on how narrow the filters are, losing information about the individual spectral features. For this reason, the existence of narrow-band surveys like J-PAS or PAU (Eriksen et al., 2019) is rewarding. With its 54 narrow-band filters, J-PAS enables to increase the precision producing full low-resolution spectra, which cover a wide range of visible wavelengths. While automatic deep learning techniques, such as convolutional neural networks, help us to deal more easily with the large amounts of data to come from current and future large-scale imaging surveys, J-PAS pseudo-spectra are useful to resolve broad spectral features of the observed objects, faster than those obtained with spectroscopic follow-ups. We expect a broad range of astrophysical studies to be carried out by the combination of these two ingredients, producing significant scientific results. Even though neural networks are really powerful on their own, we have seen that a suitable pre-processing of the data is important. In this sense, the combination of concrete

adaptive filtering techniques, like the MMF, that optimize the detection of sources on a background, and convolutional neural networks would also be an interesting project to develop in the future.

Focusing on gravitational lensing, it is a phenomenon involving two objects separated in redshift, and with different spectral behaviors, in which usually faint blue arc-like structures may be seen around luminous red galaxies. In this work we have confirmed that having observations at different wavelengths yields much better results when searching for strong lenses. Therefore, the large number of photometric bands in J-PAS can provide a clear sight for detecting lensing systems, distinguishing the source from the lens due to their spectral differences, and identifying multiple images of the source owing to the spectral similarity in the observed images. One of the main goals of J-PAS is measuring accurate photometric redshifts for galaxies up to $z \sim 1$, which significantly facilitates the discovery of lensed systems so they can be studied later by this or other surveys. The more observational data are available, the better strong lensing analyses can be performed. And this is important since lensing statistics are a promising way of deriving constraints on the cosmological parameters. Thanks to automatic detection techniques, like convolutional neural networks, and large-scale surveys, such as Euclid and LSST, which expect to observe $\sim 10^5$ strong lens systems, the search and analysis of strong lensing events is a promising research field.

Appendices

Appendix A: Derivation of the MMF formula

In this appendix, the derivation of the two-dimensional *matched multifilter* (MMF) is explained in detail. We shall begin, as we did in the derivation of the *matched filter* in Sec. 1.1, by assuming a set of images at N different frequencies observing the same area of the sky. We also assume that only one point source is within the field of view and that it is placed at the coordinate origin of the images. Therefore, the overall signal y_ν in these images is given by Eq. (1.10), and the signal s_ν from the point source is given by Eq. (1.11). Multifrequency methods take into account the statistical correlation of the noise between the different frequencies and the frequency dependence of the sources. The background noise n_ν is modeled as a homogeneous and isotropic random field with average value equal to zero, and its cross-power spectrum $\mathbf{P} = (P_{\nu_1\nu_2})$ between the noises at any two frequencies, ν_1 and ν_2, is defined in Eq. (1.28).

According to the convolution theorem declared in Eq. (1.5) and the convention of the Fourier transform chosen in Eq. (1.6), the filtered maps ω_{ψ_ν} obtained after applying a set of N linear time-invariant filters ψ_ν to the image signal from Eq. (1.11) are given by:

$$\omega_{\psi_\nu}(\mathbf{b}) = \int y_\nu(\mathbf{x})\psi_\nu(\mathbf{x};\mathbf{b})\, d\mathbf{x} = \int y_\nu(\mathbf{q})\psi_\nu(\mathbf{q})e^{-i\mathbf{q}\cdot\mathbf{b}}\, d\mathbf{q} \tag{1}$$

where $\mathbf{b}$ denotes a translation from the origin of the images, $\omega_{\psi_\nu}(\mathbf{b})$ are the filtered images in the position $\mathbf{b}$ at the corresponding frequency ν, and $y_\nu(\mathbf{q})$ and $\psi_\nu(\mathbf{q})$ are the Fourier transforms of $y_\nu(\mathbf{x})$ and $\psi_\nu(\mathbf{x})$, respectively. Therefore, each map y_ν is filtered with a linear filter ψ_ν, resulting in a filtered image ω_{ψ_ν}. The total filtered map is the sum of each one of these individual maps:

$$\omega_\psi(\mathbf{b}) = \sum_\nu \omega_{\psi_\nu}(\mathbf{b}) \tag{2}$$

and its variance is defined as

$$\sigma^2_{\omega_\psi} = \langle \omega_\psi(\mathbf{b})^2 \rangle - \langle \omega_\psi(\mathbf{b}) \rangle^2 = 2\pi \sum_{\nu_1 \nu_2} \int \psi_{\nu_1} P_{\nu_1 \nu_2} \psi_{\nu_2} \, q \, dq \tag{3}$$

Summing up, in the first step, each individual frequency image is filtered with a linear filter, and in the second step all the resulting filtered maps are combined into a single filtered image where the signal $s(x)$ has been boosted while the noise $n(x)$ has been significantly removed. Note also that this combination is completely general, since any summation coefficients different than one are absorbed in the definition of the filters ψ_ν. The only loose end left is to find the filters ψ_ν for which the total filtered field is optimal in the detection of point sources. The two necessary requirements to guarantee that the filtered field is optimal for the detection of point sources in this multifrequency approach are the same as those, (1) and (2), explained in Sec. 1.1.

In first place, the total filtered map must be an unbiased estimator, on average, of the amplitude of the source, that is, $\langle \omega_\psi(0) \rangle = A$. Therefore, the filters ψ_ν must be unbiased. In second place, the variance $\sigma^2_{\omega_\psi}$ of the total filtered map $\omega_\psi(\mathbf{b})$ has to be as small as possible, that is, the map has to be an efficient estimator of the amplitude of the source. Therefore, the filters ψ_ν must be maximum efficiency estimators. From now on, we will follow an analogous procedure to the derivation of the MF with the difference that several frequencies are being considered. Since the noise has zero mean, $\langle n(\mathbf{x}) \rangle = 0$, the first requirement can be mathematically enunciated as

$$2\pi \sum_\nu f_\nu \int \psi_\nu f_\nu \tau_\nu \, q \, dq = 1 \tag{4}$$

Given that the noise n is the only contribution to the variance $\sigma^2_{\omega_\psi}$, because the signal component does not change, we have to minimize Eq. (3) while meeting Eq. (4) at the same time. In order to achieve this, we have to resort to the method of Lagrange multipliers. The Lagrangian constructed assuming two frequency channels, ν_1 and ν_2 is the following

$$\begin{aligned} \mathcal{L} &= \sigma^2_{\omega_\psi}(\psi_{\nu_1}, \psi_{\nu_2}) + \lambda \Big[2\pi \sum_\nu f_\nu \int_0^\infty \psi_\nu(q) \tau_\nu(q) \, q \, dq - 1 \Big] \\ &= 2\pi \sum_{\nu_1 \nu_2} \int_0^\infty \psi_{\nu_1}(q) P_{\nu_1 \nu_2}(q) \psi_{\nu_2}(q) \, q \, dq + \lambda \Big[2\pi \sum_\nu f_\nu \int_0^\infty \psi_\nu(q) \tau_\nu(q) \, q \, dq - 1 \Big] \end{aligned} \tag{5}$$

We now calculate the derivatives of this Lagrangian respect to the two filters involved and equal the results to zero:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \psi_1} &= 4\pi \sum_i \int \psi_i P_{1i}\, q\, dq + \lambda f_1 \int \tau_1\, q\, dq = 0 \\ \frac{\partial \mathcal{L}}{\partial \psi_2} &= 4\pi \sum_j \int \psi_j P_{2j}\, q\, dq + \lambda f_2 \int \tau_2\, q\, dq = 0\end{aligned} \tag{6}$$

where i and j goes from 1 to 2. ψ_1 and τ_1 are the filter and the source profile at the frequency ν_1, and ψ_2 and τ_2 are the same for the frequency channel ν_2. These expressions can be rewritten as the following system of two equations:

$$\begin{aligned}2\sum_i P_{1i}\psi_i + \lambda f_1 \tau_1 = 0 \\ 2\sum_i P_{2i}\psi_i + \lambda f_2 \tau_2 = 0\end{aligned} \tag{7}$$

The simplified matrix form of this system of equations is

$$2\mathbf{P}\mathbf{\Psi} = -\lambda \mathbf{F} \tag{8}$$

where $\mathbf{\Psi} = [\psi_\nu]$ is the column vector of the filters, $\mathbf{F} = [f_\nu \tau_\nu]$ is a column vector containing the frequency dependence f_ν and the source profile τ_ν at each frequency ν, and $\mathbf{P}^{-1}$ is the inverse matrix of the cross-power spectrum $\mathbf{P}$. From the equation above is easy to obtain the formula for the two-dimensional *matched multifilter*:

$$\mathbf{\Psi}(q) = \alpha \mathbf{P}^{-1}\mathbf{F} \tag{9}$$

where α = -$\lambda/2$ is a normalization factor that preserves the source amplitude after filtering. Its expression can be obtained by substituting the MMF formula in Eq. (4):

$$\begin{aligned}2\pi \int \mathbf{F}^t \alpha \mathbf{P}^{-1}\mathbf{F}\, q\, dq = 1 \\ \frac{1}{2\pi} = \alpha \int \mathbf{F}^t \mathbf{P}^{-1}\mathbf{F}\, q\, dq \\ \alpha^{-1} = \int \mathbf{F}^t \mathbf{P}^{-1}\mathbf{F} dq\end{aligned} \tag{10}$$

Finally, we can obtain the expression for the filtered map $\sigma^2_{\omega_\psi}$ from Eq. (3):

$$\sigma^2_{\omega_\psi} = 2\pi \sum_{\nu_1\nu_2} \int \psi_{\nu_1} P_{\nu_1\nu_2} \psi_{\nu_2} = \sum_{\nu_1\nu_2} \psi_{\nu_1} P_{\nu_1\nu_2} \psi_{\nu_2}\, dq = \int \mathbf{\Psi}^t \mathbf{P} \mathbf{\Psi}\, dq \tag{11}$$

Appendix B: Selected high redshift sources

We present here two tables: the first one with the 370 high-z H-ATLAS sources from the robust high-z sample, and the second one containing the 237 high-z H-ATLAS sources from the 500 μm-riser sample. The column 'reliability' indicates if the photometry estimated using the MMF and the SPIRE channels is also consistent when incorporating the PACS bands: consistent (0), small worsening (1), not consistent (2), or PACS photometry not available (-1).

No.	H-ATLAS name	α (deg)	δ (deg)	$z_{MMF}^{unbiased}$	S_{MMF} (mJy/beam)			Reliability
					250 μm	350 μm	500 μm	
1	HATLAS J083051.0+013225	127.712684	1.540284	3.15 ± 0.34	230 ± 5	287 ± 7	265 ± 6	-1
2	HATLAS J083218.0+002527	128.075027	0.424236	1.08 ± 0.26	119 ± 7	80 ± 5	37 ± 2	-1
3	HATLAS J083344.9+000109	128.437396	0.019282	3.16 ± 0.34	58 ± 5	72 ± 6	67 ± 6	2
4	HATLAS J083726.1+015641	129.358797	1.944871	1.71 ± 0.27	118 ± 8	103 ± 7	61 ± 4	1
5	HATLAS J083904.5+022848	129.768948	2.480177	2.54 ± 0.3	48 ± 7	53 ± 7	41 ± 6	-1
6	HATLAS J083929.4+023536	129.872679	2.593495	1.66 ± 0.27	122 ± 8	104 ± 7	60 ± 4	-1
7	HATLAS J084010.0+014336	130.04181	1.726677	2.71 ± 0.31	40 ± 6	46 ± 7	38 ± 6	-1
8	HATLAS J084055.2+001819	130.230288	0.305532	1.37 ± 0.26	113 ± 7	86 ± 6	44 ± 3	1
9	HATLAS J084306.5+012342	130.777236	1.395141	3.15 ± 0.34	45 ± 5	56 ± 7	51 ± 6	-1
10	HATLAS J084615.9+012004	131.566339	1.334708	1.49 ± 0.27	114 ± 6	91 ± 5	49 ± 3	0
11	HATLAS J084641.1+002738	131.671655	0.460763	2.42 ± 0.3	54 ± 5	59 ± 6	44 ± 4	0
12	HATLAS J084658.7+021112	131.744841	2.186738	0.99 ± 0.27	205 ± 8	132 ± 5	59 ± 2	-1
13	HATLAS J084723.8+015430	131.849524	1.908484	2.34 ± 0.3	58 ± 7	61 ± 7	45 ± 5	-1
14	HATLAS J084740.8+002421	131.920354	0.405969	1.85 ± 0.28	56 ± 6	51 ± 6	32 ± 3	1
15	HATLAS J084859.1-015438	132.246479	-1.910578	2.91 ± 0.32	64 ± 4	77 ± 4	66 ± 4	-1
16	HATLAS J085015.2+005658	132.563354	0.949446	2.37 ± 0.3	83 ± 7	89 ± 7	65 ± 5	0
17	HATLAS J085034.0+003231	132.641952	0.541958	2.82 ± 0.32	46 ± 6	54 ± 7	45 ± 6	0
18	HATLAS J085111.7+004934	132.79885	0.826131	1.77 ± 0.27	115 ± 6	103 ± 6	61 ± 3	1
19	HATLAS J085133.1+014940	132.888113	1.828026	2.58 ± 0.31	43 ± 6	48 ± 7	38 ± 5	-1
20	HATLAS J085144.1+014214	132.933947	1.703908	1.35 ± 0.26	129 ± 9	98 ± 7	50 ± 3	1
21	HATLAS J085229.0+010217	133.12107	1.03812	3.02 ± 0.33	46 ± 6	56 ± 7	50 ± 6	-1
22	HATLAS J085250.9-010000	133.21219	-1.0002	1.13 ± 0.26	135 ± 8	93 ± 6	44 ± 3	0
23	HATLAS J085326.3+001146	133.35993	0.19621	2.81 ± 0.32	48 ± 7	57 ± 8	48 ± 7	0
24	HATLAS J085339.7+024216	133.41574	2.70447	1.79 ± 0.27	73 ± 7	66 ± 6	40 ± 4	-1
25	HATLAS J085341.1-003359	133.42126	-0.56665	1.47 ± 0.27	111 ± 7	88 ± 6	47 ± 3	1
26	HATLAS J085501.6+003252	133.75674	0.548	2.11 ± 0.28	48 ± 6	48 ± 6	33 ± 4	2
27	HATLAS J085521.1-003603	133.83824	-0.60105	2.9 ± 0.32	74 ± 6	89 ± 7	77 ± 6	-1
28	HATLAS J085625.4+023641	134.10592	2.61164	2.94 ± 0.33	40 ± 5	48 ± 6	42 ± 5	-1
29	HATLAS J085732.9-003320	134.38744	-0.55562	2.3 ± 0.29	45 ± 6	47 ± 6	34 ± 5	-1
30	HATLAS J085855.4+013729	134.73117	1.6249	2.37 ± 0.3	71 ± 6	76 ± 7	56 ± 5	1
31	HATLAS J085924.6+001256	134.85286	0.21562	2.77 ± 0.32	30 ± 5	35 ± 6	29 ± 5	-1
32	HATLAS J090033.8+001958	135.1412	0.33286	1.5 ± 0.27	102 ± 6	82 ± 5	45 ± 3	1
33	HATLAS J090038.0-003522	135.15864	-0.58972	1.53 ± 0.27	67 ± 6	54 ± 5	30 ± 3	1
34	HATLAS J090146.5-002054	135.4439	-0.34857	3.29 ± 0.35	37 ± 4	48 ± 5	45 ± 5	-1

Table B.1: 370 H-ATLAS sources that form our robust high-z sample.

No.	H-ATLAS name	α (deg)	δ (deg)	$z_{MMF}^{unbiased}$	S_{MMF} (mJy/beam)			Reliability
					250 μm	350 μm	500 μm	
35	HATLAS J090302.9-014127	135.76235	-1.69088	2.14 ± 0.29	313 ± 6	314 ± 6	215 ± 4	0
36	HATLAS J090311.6+003907	135.79868	0.65206	3.17 ± 0.34	139 ± 5	174 ± 6	162 ± 6	2
37	HATLAS J090318.6+012125	135.82783	1.35713	2.16 ± 0.29	67 ± 7	67 ± 7	46 ± 5	-1
38	HATLAS J090449.7+014858	136.2073	1.81633	2.19 ± 0.29	55 ± 7	56 ± 7	39 ± 5	2
39	HATLAS J090517.2+013552	136.32207	1.59802	2.46 ± 0.3	53 ± 6	58 ± 7	44 ± 5	-1
40	HATLAS J090530.1+001609	136.37546	0.26936	2.64 ± 0.31	51 ± 5	57 ± 6	46 ± 5	-1
41	HATLAS J090740.0-004200	136.91691	-0.70018	1.24 ± 0.26	450 ± 8	325 ± 6	160 ± 3	0
42	HATLAS J090844.2+000410	137.1844	0.06967	2.21 ± 0.29	54 ± 6	55 ± 6	39 ± 4	1
43	HATLAS J090930.4+002226	137.37702	0.37391	2.16 ± 0.29	55 ± 6	55 ± 6	38 ± 4	-1
44	HATLAS J090957.6-003620	137.49034	-0.60562	2.03 ± 0.28	102 ± 6	99 ± 6	65 ± 4	0
45	HATLAS J091023.0+014023	137.59615	1.67324	1.59 ± 0.27	82 ± 7	68 ± 6	38 ± 3	1
46	HATLAS J091043.0-000322	137.67949	-0.0562	1.84 ± 0.27	388 ± 7	354 ± 6	218 ± 4	0
47	HATLAS J091238.0+020048	138.15844	2.01359	2.95 ± 0.33	74 ± 6	89 ± 7	78 ± 6	-1
48	HATLAS J091304.9-005344	138.27077	-0.89563	2.54 ± 0.3	123 ± 5	136 ± 6	105 ± 4	-1
49	HATLAS J091331.4-003644	138.38089	-0.61234	1.49 ± 0.27	163 ± 7	130 ± 6	70 ± 3	1
50	HATLAS J091341.4-004343	138.42288	-0.72877	1.24 ± 0.26	145 ± 7	105 ± 5	51 ± 3	1
51	HATLAS J091350.1+014543	138.45899	1.76219	1.97 ± 0.28	93 ± 7	88 ± 7	57 ± 4	1
52	HATLAS J091457.6-003215	138.74022	-0.53761	2.53 ± 0.3	42 ± 6	47 ± 6	36 ± 5	0
53	HATLAS J091654.7+020325	139.22829	2.05695	2.08 ± 0.28	88 ± 7	87 ± 7	58 ± 5	0
54	HATLAS J091804.9+005812	139.52063	0.97017	3.13 ± 0.34	39 ± 5	48 ± 6	44 ± 6	-1
55	HATLAS J091840.8+023048	139.67015	2.51348	2.96 ± 0.33	109 ± 6	132 ± 7	116 ± 6	-1
56	HATLAS J092041.4+025254	140.1728	2.88192	2.9 ± 0.32	60 ± 5	72 ± 6	62 ± 5	-1
57	HATLAS J092110.6-002837	140.29421	-0.47712	1.08 ± 0.26	187 ± 7	125 ± 5	58 ± 2	0
58	HATLAS J092135.6+000131	140.39873	0.02553	2.09 ± 0.28	130 ± 7	129 ± 7	86 ± 4	1
59	HATLAS J092348.4+011907	140.95206	1.31888	1.51 ± 0.27	123 ± 8	100 ± 7	54 ± 4	2
60	HATLAS J092633.8-001801	141.64114	-0.30035	1.17 ± 0.26	124 ± 8	86 ± 6	41 ± 3	0
61	HATLAS J113245.7+003427	173.19058	0.57429	3.09 ± 0.33	52 ± 5	64 ± 6	59 ± 6	0
62	HATLAS J113453.6-005232	173.72346	-0.87559	1.85 ± 0.28	115 ± 7	105 ± 6	65 ± 4	1
63	HATLAS J113526.2-014606	173.85958	-1.76854	2.26 ± 0.29	260 ± 6	269 ± 6	191 ± 4	0
64	HATLAS J113526.4-011247	173.8602	-1.21325	2.04 ± 0.28	53 ± 7	51 ± 6	34 ± 4	-1
65	HATLAS J113833.0+004950	174.63769	0.83064	1.42 ± 0.27	140 ± 7	109 ± 5	58 ± 3	2
66	HATLAS J113924.2-015258	174.85095	-1.88295	2.01 ± 0.28	62 ± 8	60 ± 7	39 ± 5	1
67	HATLAS J113952.6+002604	174.96954	0.43468	2.23 ± 0.29	87 ± 6	89 ± 6	63 ± 4	1
68	HATLAS J114223.5-000033	175.59825	-0.00922	1.24 ± 0.26	126 ± 7	91 ± 5	45 ± 3	1
69	HATLAS J114227.1-015311	175.61324	-1.88666	1.54 ± 0.27	160 ± 7	131 ± 6	72 ± 3	-1
70	HATLAS J114229.0-003048	175.62116	-0.51342	2.68 ± 0.31	53 ± 6	61 ± 6	49 ± 5	2
71	HATLAS J114512.9-002446	176.30379	-0.4128	2.58 ± 0.31	31 ± 4	34 ± 5	27 ± 4	-1
72	HATLAS J114637.9-001132	176.65816	-0.19226	2.56 ± 0.31	289 ± 6	322 ± 6	252 ± 5	1
73	HATLAS J114753.1-005857	176.97166	-0.9826	2.96 ± 0.33	70 ± 5	85 ± 6	75 ± 6	2
74	HATLAS J114831.1-005312	177.1298	-0.88676	1.68 ± 0.27	123 ± 7	106 ± 6	62 ± 3	1
75	HATLAS J115101.7-020024	177.75721	-2.0068	1.9 ± 0.28	165 ± 8	153 ± 7	96 ± 5	1
76	HATLAS J115112.2-012637	177.80104	-1.4437	1.94 ± 0.28	128 ± 7	121 ± 7	77 ± 4	-1
77	HATLAS J115157.1-015046	177.98808	-1.84622	1.18 ± 0.26	125 ± 8	88 ± 5	42 ± 3	1
78	HATLAS J115303.4-011835	178.26447	-1.30988	2.28 ± 0.29	73 ± 6	76 ± 6	54 ± 4	0
79	HATLAS J115453.7-001619	178.72377	-0.27197	1.73 ± 0.27	87 ± 7	77 ± 6	45 ± 4	0
80	HATLAS J115518.1-021147	178.82556	-2.19666	1.61 ± 0.27	76 ± 8	64 ± 6	36 ± 4	-1
81	HATLAS J115540.8-005529	178.92015	-0.92491	1.87 ± 0.28	77 ± 7	71 ± 6	44 ± 4	2
82	HATLAS J115752.7-015742	179.46959	-1.96172	2.77 ± 0.32	37 ± 6	43 ± 6	36 ± 5	-1
83	HATLAS J115819.5-000127	179.58164	-0.02436	2.32 ± 0.29	55 ± 6	58 ± 6	42 ± 5	1
84	HATLAS J115850.5+005422	179.71068	0.90632	1.92 ± 0.28	50 ± 6	47 ± 6	30 ± 4	2
85	HATLAS J115922.5-010131	179.84384	-1.02528	2.42 ± 0.3	54 ± 6	58 ± 7	43 ± 5	2
86	HATLAS J120004.5-000227	180.01916	-0.04088	2.21 ± 0.29	46 ± 6	47 ± 6	33 ± 4	-1
87	HATLAS J120042.2+004322	180.17602	0.72296	2.75 ± 0.32	53 ± 6	61 ± 7	51 ± 5	-1
88	HATLAS J120127.6-014043	180.36524	-1.67888	4.06 ± 0.38	62 ± 3	91 ± 4	102 ± 5	2
89	HATLAS J120203.4-004331	180.5144	-0.72553	1.8 ± 0.27	78 ± 7	70 ± 6	42 ± 4	-1
90	HATLAS J120218.2-005922	180.57621	-0.98961	2.12 ± 0.28	61 ± 7	61 ± 7	41 ± 5	2
91	HATLAS J120319.1-011253	180.82963	-1.21497	3.02 ± 0.33	106 ± 6	129 ± 7	115 ± 6	-1
92	HATLAS J120435.9-014018	181.14969	-1.67174	2.27 ± 0.29	56 ± 6	58 ± 7	42 ± 5	-1
93	HATLAS J120604.9-020857	181.52057	-2.14942	2.42 ± 0.3	46 ± 6	49 ± 7	37 ± 5	-1
94	HATLAS J120656.4+012641	181.73502	1.44487	2.12 ± 0.28	105 ± 7	105 ± 7	71 ± 5	0
95	HATLAS J120709.2-014702	181.78856	-1.78415	2.03 ± 0.28	137 ± 7	132 ± 7	87 ± 4	2
96	HATLAS J120728.5+001557	181.86911	0.26584	2.51 ± 0.3	57 ± 7	63 ± 7	49 ± 6	-1
97	HATLAS J120753.6-011743	181.97342	-1.29552	3.34 ± 0.35	32 ± 5	41 ± 6	40 ± 6	-1
98	HATLAS J120932.7+002517	182.3866	0.42152	1.51 ± 0.27	101 ± 8	82 ± 7	44 ± 4	1
99	HATLAS J121001.9-012953	182.50821	-1.49812	1.36 ± 0.26	85 ± 9	65 ± 7	33 ± 3	1
100	HATLAS J121044.2+011848	182.68424	1.31345	1.15 ± 0.26	129 ± 7	90 ± 5	43 ± 2	-1
101	HATLAS J121218.8+010826	183.0784	1.14064	1.74 ± 0.27	54 ± 7	48 ± 6	28 ± 4	-1

Table B.1 (cont.)

No.	H-ATLAS name	α (deg)	δ (deg)	$z_{MMF}^{unbiased}$	S_{MMF} (mJy/beam)			Reliability
					250 μm	350 μm	500 μm	
102	HATLAS J121301.5-004922	183.25663	-0.82294	2.26 ± 0.29	118 ± 5	122 ± 5	87 ± 4	-1
103	HATLAS J121334.9-020323	183.39555	-2.05642	2.36 ± 0.3	170 ± 6	181 ± 6	133 ± 5	0
104	HATLAS J121542.7-005220	183.92815	-0.87229	2.35 ± 0.3	122 ± 6	129 ± 7	95 ± 5	-1
105	HATLAS J121622.8-000707	184.09513	-0.1187	2.47 ± 0.3	67 ± 6	74 ± 7	56 ± 5	0
106	HATLAS J121628.8+004153	184.12041	0.69828	2.69 ± 0.31	46 ± 6	52 ± 7	42 ± 5	-1
107	HATLAS J121741.1-011558	184.42134	-1.26633	2.06 ± 0.28	52 ± 7	51 ± 7	34 ± 5	0
108	HATLAS J121915.9+000634	184.8166	0.10965	1.8 ± 0.27	75 ± 8	67 ± 7	41 ± 4	2
109	HATLAS J121933.2-005601	184.88859	-0.93379	2.2 ± 0.29	48 ± 7	49 ± 7	34 ± 5	1
110	HATLAS J122028.2+003530	185.11783	0.59176	2.3 ± 0.29	58 ± 7	61 ± 7	44 ± 5	0
111	HATLAS J122125.7-014641	185.35747	-1.77819	2.79 ± 0.32	41 ± 6	48 ± 7	40 ± 6	-1
112	HATLAS J122127.9+005026	185.36626	0.84058	1.97 ± 0.28	116 ± 7	111 ± 7	71 ± 4	1
113	HATLAS J122134.0-004919	185.39173	-0.82196	2.34 ± 0.29	50 ± 7	53 ± 7	39 ± 5	-1
114	HATLAS J122201.5-011354	185.50653	-1.23177	2.43 ± 0.3	56 ± 7	60 ± 7	45 ± 6	1
115	HATLAS J122205.8+013232	185.52438	1.54247	1.79 ± 0.27	65 ± 7	58 ± 6	35 ± 4	0
116	HATLAS J122247.3+001906	185.69734	0.31843	3.03 ± 0.33	71 ± 5	87 ± 6	78 ± 6	-1
117	HATLAS J122335.1+004307	185.89649	0.71884	1.81 ± 0.27	84 ± 6	76 ± 6	46 ± 4	2
118	HATLAS J122342.4-001103	185.92673	-0.1842	1.74 ± 0.27	84 ± 7	74 ± 6	44 ± 4	1
119	HATLAS J140223.0+010213	210.59604	1.03696	2.14 ± 0.29	101 ± 7	102 ± 7	70 ± 5	1
120	HATLAS J140701.1-004402	211.75489	-0.73406	1.98 ± 0.28	64 ± 7	61 ± 6	40 ± 4	-1
121	HATLAS J140727.9+013759	211.8664	1.63332	2.74 ± 0.31	62 ± 6	72 ± 7	59 ± 6	-1
122	HATLAS J141010.2-002700	212.54269	-0.45009	1.87 ± 0.28	71 ± 7	65 ± 7	41 ± 4	1
123	HATLAS J141038.9-014816	212.66209	-1.8046	1.89 ± 0.28	91 ± 6	84 ± 6	53 ± 3	-1
124	HATLAS J141110.3-001715	212.7931	-0.28775	2.12 ± 0.29	51 ± 7	51 ± 7	35 ± 5	2
125	HATLAS J141149.6-005632	212.95704	-0.94239	1.67 ± 0.27	111 ± 8	96 ± 7	55 ± 4	1
126	HATLAS J141246.4+022948	213.19373	2.49675	3.18 ± 0.34	38 ± 5	48 ± 6	44 ± 6	0
127	HATLAS J141327.0+012305	213.36252	1.38473	1.8 ± 0.27	58 ± 7	53 ± 6	32 ± 4	0
128	HATLAS J141328.7-012251	213.3698	-1.38108	1.87 ± 0.28	76 ± 7	70 ± 6	44 ± 4	-1
129	HATLAS J141346.1-000928	213.44232	-0.158	1.47 ± 0.27	123 ± 8	98 ± 6	52 ± 3	0
130	HATLAS J141351.9-000026	213.46663	-0.00745	2.73 ± 0.31	166 ± 5	192 ± 6	158 ± 5	0
131	HATLAS J141409.7+022541	213.54075	2.42811	2.45 ± 0.3	54 ± 7	59 ± 7	44 ± 5	1
132	HATLAS J141425.0+013039	213.60457	1.51091	2.93 ± 0.33	38 ± 5	46 ± 6	40 ± 5	2
133	HATLAS J141528.0+005204	213.86674	0.86779	1.28 ± 0.26	130 ± 7	96 ± 5	48 ± 3	1
134	HATLAS J141550.4+012750	213.96003	1.46401	2.77 ± 0.32	36 ± 5	42 ± 6	35 ± 5	1
135	HATLAS J141550.8+001546	213.96188	0.2628	2.53 ± 0.3	64 ± 6	71 ± 6	55 ± 5	1
136	HATLAS J141559.5+013522	213.99825	1.58968	2.72 ± 0.31	46 ± 5	54 ± 6	44 ± 5	1
137	HATLAS J141605.5+011828	214.02308	1.30802	2.54 ± 0.3	55 ± 6	61 ± 6	47 ± 5	1
138	HATLAS J141718.2+005554	214.32618	0.9319	2.21 ± 0.29	57 ± 6	58 ± 6	41 ± 4	-1
139	HATLAS J141729.9+001637	214.37494	0.27719	1.96 ± 0.28	42 ± 7	39 ± 6	25 ± 4	1
140	HATLAS J141750.9+003054	214.46215	0.515	3.24 ± 0.34	35 ± 4	44 ± 5	41 ± 5	1
141	HATLAS J141805.6-002140	214.52369	-0.36132	2.26 ± 0.29	47 ± 6	48 ± 6	34 ± 5	1
142	HATLAS J141827.4-001703	214.61445	-0.28434	1.86 ± 0.28	104 ± 6	95 ± 6	59 ± 4	0
143	HATLAS J141832.9+010212	214.63749	1.03685	3.9 ± 0.38	51 ± 4	73 ± 6	80 ± 6	-1
144	HATLAS J141852.9-010138	214.72082	-1.02729	2.22 ± 0.29	46 ± 6	47 ± 7	33 ± 5	-1
145	HATLAS J141955.5-003449	214.98151	-0.58029	1.82 ± 0.27	160 ± 7	145 ± 6	88 ± 4	2
146	HATLAS J142008.9-001434	215.03741	-0.24285	2.27 ± 0.29	73 ± 6	76 ± 6	54 ± 4	-1
147	HATLAS J142033.9-014111	215.14149	-1.68641	1.65 ± 0.27	66 ± 7	56 ± 6	32 ± 3	2
148	HATLAS J142046.8-005721	215.19507	-0.95596	2.98 ± 0.33	48 ± 6	58 ± 7	51 ± 6	-1
149	HATLAS J142049.1+000555	215.20495	0.09878	2.49 ± 0.3	50 ± 7	54 ± 7	42 ± 6	1
150	HATLAS J142112.4+005230	215.30175	0.87508	2.29 ± 0.29	88 ± 7	92 ± 7	66 ± 5	0
151	HATLAS J142123.2+002617	215.34704	0.43824	2.34 ± 0.29	54 ± 7	57 ± 7	42 ± 5	1
152	HATLAS J142142.1+015229	215.42571	1.87488	2.56 ± 0.31	38 ± 7	42 ± 8	33 ± 6	1
153	HATLAS J142233.9+023413	215.64142	2.57052	2.26 ± 0.29	57 ± 5	59 ± 5	42 ± 4	0
154	HATLAS J142303.2+010716	215.76366	1.12112	2.32 ± 0.29	68 ± 6	71 ± 7	52 ± 5	-1
155	HATLAS J142318.3+013913	215.82658	1.65383	2.45 ± 0.3	100 ± 7	109 ± 8	82 ± 6	0
156	HATLAS J142339.4+003242	215.91453	0.54504	2.55 ± 0.31	59 ± 6	66 ± 7	51 ± 5	-1
157	HATLAS J142413.9+022303	216.0582	2.38424	3.94 ± 0.38	117 ± 4	169 ± 6	186 ± 7	-1
158	HATLAS J142431.9-003932	216.13306	-0.65909	2.34 ± 0.3	63 ± 7	67 ± 8	49 ± 6	-1
159	HATLAS J142540.9+013151	216.42074	1.53096	3.26 ± 0.34	37 ± 5	48 ± 7	45 ± 6	-1
160	HATLAS J142554.5+012650	216.47728	1.44723	2.14 ± 0.29	52 ± 7	53 ± 7	36 ± 5	-1
161	HATLAS J142629.0+014412	216.62094	1.73685	3.31 ± 0.35	36 ± 5	47 ± 6	45 ± 6	-1
162	HATLAS J142702.9+003053	216.7624	0.51473	2.37 ± 0.3	56 ± 6	60 ± 7	44 ± 5	2
163	HATLAS J142736.9+022920	216.90386	2.48907	3.18 ± 0.34	44 ± 5	55 ± 6	51 ± 5	-1
164	HATLAS J142737.1+021335	216.90491	2.22665	2.61 ± 0.31	52 ± 6	59 ± 7	46 ± 6	-1
165	HATLAS J142751.0+004233	216.96278	0.70933	1.85 ± 0.28	122 ± 6	112 ± 6	69 ± 3	1
166	HATLAS J142810.7+002843	217.04469	0.47868	1.67 ± 0.27	113 ± 7	98 ± 6	56 ± 4	0
167	HATLAS J142824.7+013930	217.10293	1.65847	1.49 ± 0.27	101 ± 7	81 ± 6	44 ± 3	2
168	HATLAS J142859.8+015037	217.24956	1.84381	3.34 ± 0.35	33 ± 5	43 ± 6	42 ± 6	2

Table B.1 (cont.)

No.	H-ATLAS name	α (deg)	δ (deg)	$z_{MMF}^{unbiased}$	S_{MMF} (mJy/beam)			Reliability
					250 μm	350 μm	500 μm	
169	HATLAS J142935.5+012749	217.39832	1.46386	2.56 ± 0.31	40 ± 6	44 ± 6	35 ± 5	-1
170	HATLAS J143135.6+003041	217.89868	0.51158	2.39 ± 0.3	45 ± 6	48 ± 6	36 ± 4	-1
171	HATLAS J143229.4-000820	218.12257	-0.1391	2.58 ± 0.31	66 ± 5	74 ± 6	58 ± 5	2
172	HATLAS J143358.4-012718	218.49358	-1.45503	1.89 ± 0.28	63 ± 6	59 ± 6	37 ± 4	2
173	HATLAS J143441.2-014548	218.67174	-1.76335	2.35 ± 0.3	53 ± 6	56 ± 7	41 ± 5	-1
174	HATLAS J143505.9-002139	218.7746	-0.36105	2.12 ± 0.29	60 ± 7	59 ± 7	40 ± 5	-1
175	HATLAS J143512.3+004222	218.80147	0.70618	2.98 ± 0.33	39 ± 5	48 ± 7	42 ± 6	2
176	HATLAS J143512.3+021203	218.80161	2.20094	1.2 ± 0.26	119 ± 8	84 ± 6	41 ± 3	2
177	HATLAS J143530.8+013240	218.87848	1.54458	2.31 ± 0.29	49 ± 6	52 ± 7	37 ± 5	1
178	HATLAS J143707.1+011050	219.27987	1.18082	3.14 ± 0.34	31 ± 4	38 ± 5	35 ± 5	-1
179	HATLAS J143735.5-013804	219.3982	-1.63455	2.04 ± 0.28	84 ± 6	82 ± 6	54 ± 4	-1
180	HATLAS J143841.2+013110	219.6718	1.51948	3.1 ± 0.33	36 ± 5	45 ± 7	41 ± 6	1
181	HATLAS J143845.8+013503	219.69125	1.58435	1.73 ± 0.27	79 ± 7	69 ± 6	41 ± 4	-1
182	HATLAS J143852.1+021130	219.71739	2.1917	1.03 ± 0.26	120 ± 7	79 ± 5	36 ± 2	1
183	HATLAS J143853.4+001904	219.72263	0.31781	2.41 ± 0.3	53 ± 7	57 ± 8	43 ± 6	1
184	HATLAS J143936.0+003839	219.90036	0.64437	1.7 ± 0.27	110 ± 6	96 ± 5	56 ± 3	2
185	HATLAS J144042.1+003316	220.17561	0.55466	2.26 ± 0.29	59 ± 6	61 ± 6	43 ± 4	0
186	HATLAS J144154.3-001345	220.47638	-0.22941	1.68 ± 0.27	120 ± 7	104 ± 6	60 ± 3	0
187	HATLAS J144201.7+003449	220.50743	0.58036	1.92 ± 0.28	83 ± 7	78 ± 6	49 ± 4	1
188	HATLAS J144243.4+015504	220.68089	1.9179	2.19 ± 0.29	120 ± 7	122 ± 7	84 ± 5	1
189	HATLAS J144334.3-003034	220.89297	-0.50953	3.27 ± 0.34	64 ± 4	82 ± 5	78 ± 5	-1
190	HATLAS J144445.3-000910	221.18892	-0.15301	1.68 ± 0.27	55 ± 8	47 ± 7	27 ± 4	0
191	HATLAS J144521.7+005504	221.34062	0.91801	2.62 ± 0.31	36 ± 6	41 ± 7	33 ± 5	2
192	HATLAS J144524.8+000002	221.3534	0.00058	2.92 ± 0.33	31 ± 5	38 ± 6	33 ± 5	-1
193	HATLAS J144533.8+010556	221.39124	1.09906	1.35 ± 0.26	168 ± 7	127 ± 5	65 ± 3	2
194	HATLAS J144556.1-004853	221.48377	-0.81481	2.43 ± 0.3	112 ± 6	121 ± 7	91 ± 5	-1
195	HATLAS J144626.3+011400	221.60991	1.23354	2.34 ± 0.29	50 ± 6	53 ± 7	39 ± 5	1
196	HATLAS J144828.6+020947	222.11918	2.16307	2.79 ± 0.32	58 ± 6	68 ± 7	57 ± 6	0
197	HATLAS J145138.5-012819	222.91073	-1.472	1.55 ± 0.27	117 ± 6	96 ± 5	53 ± 3	-1
198	HATLAS J145337.2+000407	223.4052	0.06887	2.55 ± 0.31	63 ± 7	70 ± 7	55 ± 6	0
199	HATLAS J145342.8-002345	223.42844	-0.39601	2.42 ± 0.3	61 ± 6	66 ± 7	49 ± 5	0
200	HATLAS J145420.6-005203	223.58616	-0.86764	2.58 ± 0.31	29 ± 6	32 ± 6	25 ± 5	-1
201	HATLAS J145653.4-000720	224.2226	-0.12243	1.58 ± 0.27	62 ± 7	51 ± 6	29 ± 3	0
202	HATLAS J000007.5-334100	0.03119	-33.68326	2.25 ± 0.29	136 ± 7	141 ± 7	100 ± 5	-1
203	HATLAS J000525.0-331342	1.35423	-33.22837	1.27 ± 0.26	136 ± 8	99 ± 6	50 ± 3	1
204	HATLAS J000622.1-300508	1.59225	-30.0856	1.35 ± 0.26	145 ± 7	110 ± 6	56 ± 3	1
205	HATLAS J000631.2-312238	1.63017	-31.37732	1.01 ± 0.26	149 ± 8	97 ± 5	44 ± 2	1
206	HATLAS J000722.2-352015	1.84235	-35.33737	1.61 ± 0.27	210 ± 7	177 ± 6	100 ± 3	1
207	HATLAS J000806.8-351205	2.02835	-35.20138	3.18 ± 0.34	75 ± 5	95 ± 6	88 ± 5	-1
208	HATLAS J000912.7-300807	2.30293	-30.13529	1.3 ± 0.26	322 ± 9	238 ± 6	120 ± 3	0
209	HATLAS J001010.5-360237	2.54381	-36.04354	2.03 ± 0.28	152 ± 7	147 ± 7	97 ± 5	-1
210	HATLAS J001148.2-312431	2.95101	-31.40849	2.1 ± 0.28	107 ± 7	106 ± 7	71 ± 5	2
211	HATLAS J001334.2-324432	3.39265	-32.74231	2.06 ± 0.28	117 ± 7	115 ± 7	76 ± 5	1
212	HATLAS J001725.2-295152	4.35502	-29.86437	1.89 ± 0.28	140 ± 7	130 ± 7	81 ± 4	0
213	HATLAS J001802.2-313505	4.50901	-31.58466	1.85 ± 0.28	126 ± 7	116 ± 7	71 ± 4	0
214	HATLAS J002533.6-333826	6.38992	-33.64055	2.33 ± 0.29	116 ± 6	122 ± 7	89 ± 5	1
215	HATLAS J002624.8-341738	6.60345	-34.29384	2.89 ± 0.32	138 ± 5	165 ± 6	142 ± 5	-1
216	HATLAS J002913.4-322500	7.30577	-32.41662	2.17 ± 0.29	100 ± 7	101 ± 7	69 ± 5	0
217	HATLAS J002953.4-322318	7.47262	-32.38842	1.75 ± 0.27	154 ± 6	136 ± 6	81 ± 3	1
218	HATLAS J003605.8-290025	9.02405	-29.00704	1.35 ± 0.26	199 ± 8	150 ± 6	77 ± 3	1
219	HATLAS J003905.5-290609	9.77295	-29.10258	1.99 ± 0.28	102 ± 6	98 ± 6	63 ± 4	1
220	HATLAS J004130.7-332921	10.37795	-33.48922	1.45 ± 0.27	132 ± 8	105 ± 6	56 ± 3	1
221	HATLAS J004736.0-272951	11.89999	-27.49738	2.71 ± 0.31	156 ± 6	180 ± 6	147 ± 5	-1
222	HATLAS J004853.3-303110	12.22192	-30.51932	2.39 ± 0.3	117 ± 6	126 ± 6	93 ± 5	0
223	HATLAS J005119.0-312321	12.82908	-31.38921	1.7 ± 0.27	123 ± 7	107 ± 6	63 ± 4	0
224	HATLAS J005132.8-301848	12.8867	-30.31335	2.21 ± 0.29	151 ± 7	155 ± 7	108 ± 5	1
225	HATLAS J005212.9-334502	13.05365	-33.75058	1.35 ± 0.26	194 ± 8	146 ± 6	75 ± 3	1
226	HATLAS J005809.8-330233	14.54096	-33.04264	1.3 ± 0.26	135 ± 6	100 ± 5	50 ± 2	1
227	HATLAS J005838.3-331143	14.65957	-33.19531	1.2 ± 0.26	124 ± 7	88 ± 5	42 ± 3	1
228	HATLAS J005952.1-303802	14.96708	-30.63388	2.11 ± 0.28	91 ± 6	91 ± 6	61 ± 4	-1
229	HATLAS J005957.0-284530	14.98731	-28.75834	1.06 ± 0.26	123 ± 7	82 ± 5	38 ± 2	2
230	HATLAS J010250.9-311723	15.71188	-31.28976	1.96 ± 0.28	246 ± 8	233 ± 7	149 ± 5	1
231	HATLAS J010753.0-305344	16.971	-30.89549	1.55 ± 0.27	130 ± 8	107 ± 6	59 ± 3	2
232	HATLAS J011424.0-333614	18.60015	-33.6038	3.97 ± 0.38	82 ± 4	119 ± 6	131 ± 7	2
233	HATLAS J011512.1-283958	18.80045	-28.66613	1.2 ± 0.26	137 ± 8	97 ± 5	47 ± 3	0
234	HATLAS J011542.0-273553	18.92486	-27.59819	1.74 ± 0.27	110 ± 7	97 ± 6	57 ± 4	-1
235	HATLAS J011823.8-274404	19.59911	-27.7344	2.37 ± 0.3	101 ± 6	108 ± 7	80 ± 5	-1

Table B.1 (cont.)

No.	H-ATLAS name	α (deg)	δ (deg)	$z_{MMF}^{unbiased}$	S_{MMF} (mJy/beam)			Reliability
					250 μm	350 μm	500 μm	
236	HATLAS J012046.5-282403	20.19358	-28.40097	3.51 ± 0.36	105 ± 4	140 ± 6	140 ± 6	-1
237	HATLAS J012407.4-281434	21.03082	-28.24284	2.45 ± 0.3	228 ± 6	248 ± 7	187 ± 5	1
238	HATLAS J012416.0-310500	21.06657	-31.08338	2.42 ± 0.3	124 ± 7	134 ± 7	100 ± 5	-1
239	HATLAS J012853.0-332719	22.22084	-33.45536	2.52 ± 0.3	97 ± 6	107 ± 6	83 ± 5	-1
240	HATLAS J013004.1-305514	22.51707	-30.92051	1.87 ± 0.28	145 ± 6	134 ± 6	83 ± 4	1
241	HATLAS J013240.0-330907	23.16656	-33.15184	2.64 ± 0.31	104 ± 6	119 ± 7	95 ± 6	-1
242	HATLAS J013524.3-330959	23.85142	-33.16643	1.84 ± 0.27	155 ± 9	141 ± 8	87 ± 5	0
243	HATLAS J013732.3-314506	24.38473	-31.75164	1.48 ± 0.27	124 ± 8	99 ± 7	53 ± 4	0
244	HATLAS J013840.5-281856	24.66875	-28.31545	4.39 ± 0.39	93 ± 4	150 ± 6	177 ± 7	-1
245	HATLAS J014849.3-331820	27.20521	-33.30547	2.33 ± 0.29	125 ± 8	132 ± 8	96 ± 6	-1
246	HATLAS J124957.4+270606	192.48913	27.1017	1.2 ± 0.26	134 ± 9	95 ± 6	46 ± 3	0
247	HATLAS J125000.2+312738	192.50064	31.46054	1.56 ± 0.27	102 ± 6	84 ± 5	47 ± 3	0
248	HATLAS J125000.4+305323	192.50182	30.8896	2.17 ± 0.29	73 ± 8	74 ± 8	51 ± 6	0
249	HATLAS J125027.9+242643	192.61607	24.44523	1.8 ± 0.27	77 ± 7	70 ± 6	42 ± 4	1
250	HATLAS J125233.5+331031	193.13953	33.17537	2.06 ± 0.28	58 ± 6	56 ± 6	37 ± 4	2
251	HATLAS J125404.7+312610	193.51968	31.43623	2.02 ± 0.28	59 ± 7	57 ± 7	38 ± 5	1
252	HATLAS J125441.2+333753	193.67185	33.63146	1.51 ± 0.27	217 ± 8	175 ± 6	95 ± 3	-1
253	HATLAS J125526.4+304930	193.86011	30.8249	1.43 ± 0.27	167 ± 8	131 ± 6	69 ± 3	0
254	HATLAS J125532.6+325638	193.88574	32.94389	1.36 ± 0.26	111 ± 7	85 ± 5	44 ± 3	2
255	HATLAS J125632.5+233627	194.13524	23.60758	3.39 ± 0.35	201 ± 5	261 ± 7	256 ± 6	1
256	HATLAS J125700.6+312507	194.25238	31.41864	1.02 ± 0.26	131 ± 8	85 ± 5	39 ± 2	1
257	HATLAS J125710.9+241533	194.29549	24.25923	1.07 ± 0.26	139 ± 8	93 ± 5	43 ± 3	0
258	HATLAS J125911.0+293843	194.79574	29.6454	1.98 ± 0.28	151 ± 7	145 ± 7	93 ± 4	0
259	HATLAS J130138.7+302808	195.41128	30.46876	2.03 ± 0.28	45 ± 7	43 ± 7	29 ± 4	1
260	HATLAS J130152.8+245012	195.46998	24.8366	2.36 ± 0.3	46 ± 7	49 ± 7	36 ± 5	1
261	HATLAS J130314.0+315646	195.80823	31.94621	1.8 ± 0.27	127 ± 7	115 ± 6	69 ± 4	1
262	HATLAS J130333.1+244643	195.8881	24.77864	2.72 ± 0.31	96 ± 6	111 ± 6	91 ± 5	2
263	HATLAS J130334.4+282029	195.89353	28.34151	2.34 ± 0.29	51 ± 7	54 ± 7	39 ± 5	-1
264	HATLAS J130349.8+351024	195.95757	35.17324	1.54 ± 0.27	98 ± 7	80 ± 6	44 ± 3	-1
265	HATLAS J130525.0+342852	196.3543	34.48122	1.85 ± 0.28	121 ± 8	111 ± 7	68 ± 4	2
266	HATLAS J130602.9+333035	196.5121	33.5098	1.23 ± 0.26	128 ± 8	92 ± 6	45 ± 3	1
267	HATLAS J131025.0+301933	197.60417	30.3257	2.91 ± 0.32	40 ± 6	48 ± 7	42 ± 6	-1
268	HATLAS J131105.4+253024	197.77244	25.50653	2.35 ± 0.3	61 ± 6	65 ± 6	47 ± 5	-1
269	HATLAS J131301.8+305942	198.25761	30.99498	1.73 ± 0.27	131 ± 8	115 ± 7	68 ± 4	1
270	HATLAS J131348.7+242456	198.45272	24.41548	2.47 ± 0.3	91 ± 6	99 ± 7	75 ± 5	0
271	HATLAS J131407.0+271208	198.52902	27.20212	0.96 ± 0.27	146 ± 7	92 ± 4	41 ± 2	1
272	HATLAS J131435.2+301802	198.64671	30.30066	2.2 ± 0.29	86 ± 8	88 ± 8	61 ± 6	0
273	HATLAS J131533.9+233254	198.89142	23.54824	1.36 ± 0.26	119 ± 7	90 ± 6	47 ± 3	2
274	HATLAS J131540.6+262322	198.91917	26.38947	3.24 ± 0.34	85 ± 5	108 ± 7	102 ± 6	0
275	HATLAS J131634.2+243457	199.14232	24.58243	2.37 ± 0.3	63 ± 6	67 ± 7	50 ± 5	2
276	HATLAS J131635.1+332406	199.14629	33.40161	3.33 ± 0.35	34 ± 4	44 ± 5	43 ± 5	-1
277	HATLAS J131715.6+322643	199.3149	32.44538	1.25 ± 0.26	79 ± 7	57 ± 5	28 ± 2	2
278	HATLAS J131820.5+230158	199.5856	23.03276	1.08 ± 0.26	154 ± 9	103 ± 6	48 ± 3	-1
279	HATLAS J132034.0+343940	200.14147	34.661	1.39 ± 0.26	120 ± 8	92 ± 6	48 ± 3	1
280	HATLAS J132111.2+265009	200.29654	26.83585	2.12 ± 0.28	99 ± 6	99 ± 6	67 ± 4	1
281	HATLAS J132123.8+265910	200.34925	26.9861	1.61 ± 0.27	114 ± 8	96 ± 6	54 ± 4	-1
282	HATLAS J132128.6+282020	200.36904	28.33888	2.17 ± 0.29	100 ± 6	101 ± 6	70 ± 4	-1
283	HATLAS J132314.3+285039	200.80957	28.84426	1.68 ± 0.27	154 ± 8	133 ± 7	77 ± 4	-1
284	HATLAS J132355.2+282319	200.97994	28.3885	2.28 ± 0.29	63 ± 6	66 ± 6	47 ± 5	1
285	HATLAS J132419.0+320752	201.07921	32.13114	3.75 ± 0.37	71 ± 4	99 ± 5	104 ± 6	2
286	HATLAS J132427.0+284450	201.11257	28.74716	2.27 ± 0.29	329 ± 7	342 ± 7	244 ± 5	2
287	HATLAS J132453.6+244909	201.22317	24.81924	1.95 ± 0.28	55 ± 7	52 ± 6	33 ± 4	-1
288	HATLAS J132504.4+311534	201.26821	31.25951	1.83 ± 0.27	233 ± 8	212 ± 7	130 ± 4	0
289	HATLAS J132621.6+231646	201.58983	23.27955	1.64 ± 0.27	68 ± 7	58 ± 6	33 ± 3	-1
290	HATLAS J132630.1+334408	201.6255	33.73548	3.94 ± 0.38	171 ± 4	247 ± 6	270 ± 7	2
291	HATLAS J132630.5+291840	201.62697	29.31112	1.32 ± 0.26	138 ± 8	103 ± 6	52 ± 3	1
292	HATLAS J132748.0+282411	201.94988	28.40301	1.44 ± 0.27	130 ± 8	102 ± 6	54 ± 3	1
293	HATLAS J132812.5+232747	202.05204	23.46302	1.85 ± 0.28	78 ± 8	71 ± 7	44 ± 4	1
294	HATLAS J132859.2+292327	202.2468	29.39074	2.85 ± 0.32	233 ± 5	275 ± 6	234 ± 5	0
295	HATLAS J132900.4+281914	202.25187	28.32064	2.32 ± 0.29	110 ± 6	116 ± 6	84 ± 4	-1
296	HATLAS J132957.3+295302	202.48869	29.88398	1.37 ± 0.26	140 ± 8	107 ± 6	55 ± 3	0
297	HATLAS J133008.6+245900	202.53579	24.98334	2.33 ± 0.29	255 ± 6	269 ± 7	196 ± 5	1
298	HATLAS J133020.9+240249	202.58712	24.04686	3.01 ± 0.33	53 ± 6	65 ± 7	58 ± 6	-1
299	HATLAS J133038.3+255129	202.65945	25.85796	2.04 ± 0.28	151 ± 7	147 ± 7	97 ± 5	0
300	HATLAS J133231.6+350843	203.13181	35.14522	1.82 ± 0.27	114 ± 7	104 ± 7	63 ± 4	2
301	HATLAS J133255.8+342208	203.23246	34.36886	1.8 ± 0.27	189 ± 8	170 ± 7	103 ± 4	-1
302	HATLAS J133312.2+271025	203.30088	27.17365	1.21 ± 0.26	145 ± 8	103 ± 6	50 ± 3	1

Table B.1 (cont.)

No.	H-ATLAS name	α (deg)	δ (deg)	$z_{MMF}^{unbiased}$	S_{MMF} (mJy/beam)			Reliability
					250 μm	350 μm	500 μm	
303	HATLAS J133413.8+260458	203.55768	26.08276	2.63 ± 0.31	127 ± 5	144 ± 6	115 ± 5	2
304	HATLAS J133440.4+353141	203.66844	35.52806	2.92 ± 0.32	69 ± 5	83 ± 5	72 ± 5	2
305	HATLAS J133534.1+341835	203.892	34.30967	2.3 ± 0.29	106 ± 6	111 ± 6	80 ± 5	1
306	HATLAS J133538.3+265742	203.90947	26.96173	2.14 ± 0.29	98 ± 6	98 ± 6	67 ± 4	-1
307	HATLAS J133543.0+300402	203.92905	30.06709	2.26 ± 0.29	130 ± 6	134 ± 7	95 ± 5	0
308	HATLAS J133622.7+343806	204.09456	34.63513	1.22 ± 0.26	190 ± 9	136 ± 6	67 ± 3	1
309	HATLAS J133650.0+291800	204.20819	29.3	1.83 ± 0.27	291 ± 7	265 ± 6	162 ± 4	0
310	HATLAS J133806.8+351530	204.52854	35.25838	1.9 ± 0.28	79 ± 7	74 ± 6	46 ± 4	-1
311	HATLAS J133834.2+294805	204.64238	29.80143	1.37 ± 0.26	135 ± 9	103 ± 7	53 ± 3	1
312	HATLAS J133846.5+255055	204.69389	25.84855	2.45 ± 0.3	151 ± 7	165 ± 7	125 ± 5	2
313	HATLAS J133856.5+304721	204.73523	30.78917	1.35 ± 0.26	154 ± 7	116 ± 5	60 ± 3	0
314	HATLAS J133905.0+340819	204.77072	34.13859	1.9 ± 0.28	129 ± 7	120 ± 7	76 ± 4	0
315	HATLAS J134132.5+263334	205.38559	26.55947	1.67 ± 0.27	98 ± 6	84 ± 5	49 ± 3	1
316	HATLAS J134158.5+292833	205.49377	29.47585	1.59 ± 0.27	186 ± 8	155 ± 6	87 ± 4	-1
317	HATLAS J134240.1+272356	205.66709	27.39888	1.84 ± 0.27	102 ± 7	93 ± 6	57 ± 4	1
318	HATLAS J134324.5+240202	205.85212	24.03402	2.15 ± 0.29	73 ± 7	73 ± 7	50 ± 5	-1
319	HATLAS J134400.4+342243	206.00159	34.37868	1.59 ± 0.27	140 ± 8	117 ± 6	66 ± 4	1
320	HATLAS J134421.9+261513	206.09107	26.25358	1.4 ± 0.27	124 ± 8	96 ± 6	50 ± 3	2
321	HATLAS J134429.5+303034	206.12285	30.50948	2.29 ± 0.29	443 ± 6	463 ± 7	332 ± 5	0
322	HATLAS J134441.5+240345	206.17285	24.06259	2.37 ± 0.3	65 ± 7	69 ± 7	51 ± 5	-1
323	HATLAS J134553.7+345802	206.47383	34.96729	1.02 ± 0.26	134 ± 8	87 ± 5	39 ± 2	0
324	HATLAS J134654.4+295659	206.72685	29.94962	1.56 ± 0.27	143 ± 8	118 ± 6	66 ± 3	1
325	HATLAS J134853.0+270011	207.22075	27.00295	2.49 ± 0.3	71 ± 6	78 ± 7	60 ± 5	-1
326	HATLAS J135020.3+265536	207.58468	26.92669	2.04 ± 0.28	148 ± 7	144 ± 7	95 ± 4	-1
327	HATLAS J135026.2+241402	207.60907	24.23384	2.78 ± 0.32	37 ± 6	43 ± 7	36 ± 6	-1
328	HATLAS J222206.2-340025	335.52571	-34.00681	1.29 ± 0.26	159 ± 8	117 ± 6	59 ± 3	2
329	HATLAS J223753.8-305828	339.47426	-30.97452	2.26 ± 0.29	121 ± 6	126 ± 6	89 ± 4	2
330	HATLAS J223829.1-304149	339.62142	-30.69684	1.5 ± 0.27	226 ± 8	182 ± 6	99 ± 3	0
331	HATLAS J223942.4-333304	339.92682	-33.55121	2.39 ± 0.3	91 ± 6	97 ± 6	72 ± 5	-1
332	HATLAS J224026.5-315155	340.11058	-31.86522	1.92 ± 0.28	102 ± 7	96 ± 6	60 ± 4	0
333	HATLAS J224435.5-324202	341.14786	-32.70061	1.66 ± 0.27	183 ± 9	157 ± 8	90 ± 5	0
334	HATLAS J224506.5-295624	341.27704	-29.94002	1.06 ± 0.26	191 ± 8	127 ± 5	58 ± 2	2
335	HATLAS J224536.3-312841	341.40112	-31.47813	1.15 ± 0.26	129 ± 8	89 ± 6	43 ± 3	1
336	HATLAS J224720.9-302206	341.837	-30.36833	1.22 ± 0.26	132 ± 8	94 ± 5	46 ± 3	2
337	HATLAS J224805.4-335820	342.02232	-33.97226	2.87 ± 0.32	108 ± 5	128 ± 6	110 ± 6	0
338	HATLAS J224932.7-323116	342.38612	-32.52122	1.81 ± 0.27	107 ± 7	97 ± 6	59 ± 4	0
339	HATLAS J225201.1-322055	343.00477	-32.34864	1.15 ± 0.26	150 ± 7	104 ± 5	49 ± 2	2
340	HATLAS J225250.7-313658	343.21138	-31.61605	2.7 ± 0.31	98 ± 5	112 ± 6	92 ± 5	0
341	HATLAS J225301.6-314442	343.25682	-31.74502	1.49 ± 0.27	127 ± 7	102 ± 6	55 ± 3	0
342	HATLAS J225324.2-323504	343.35103	-32.5845	2.05 ± 0.28	118 ± 7	115 ± 7	76 ± 4	0
343	HATLAS J225600.7-313232	344.00286	-31.54211	1.86 ± 0.28	118 ± 7	109 ± 6	67 ± 4	2
344	HATLAS J225706.7-294314	344.27786	-29.72061	1.44 ± 0.27	155 ± 9	122 ± 7	65 ± 4	2
345	HATLAS J225844.8-295125	344.68666	-29.85693	2.44 ± 0.3	156 ± 7	169 ± 7	128 ± 6	1
346	HATLAS J230815.6-343801	347.06496	-34.6337	4.03 ± 0.38	81 ± 4	119 ± 6	132 ± 6	-1
347	HATLAS J230917.7-312144	347.32394	-31.36222	2.06 ± 0.28	105 ± 7	103 ± 7	68 ± 4	1
348	HATLAS J231651.6-345845	349.2149	-34.97919	1.7 ± 0.27	118 ± 7	103 ± 6	60 ± 4	-1
349	HATLAS J232210.9-333749	350.5454	-33.63041	1.89 ± 0.28	106 ± 6	99 ± 6	62 ± 4	1
350	HATLAS J232419.8-323927	351.08254	-32.65738	2.47 ± 0.3	199 ± 5	217 ± 6	165 ± 4	0
351	HATLAS J232531.4-302236	351.38064	-30.37653	2.32 ± 0.29	191 ± 5	201 ± 6	146 ± 4	-1
352	HATLAS J232550.8-333826	351.46158	-33.64046	1.89 ± 0.28	130 ± 7	121 ± 6	75 ± 4	0
353	HATLAS J232623.0-342642	351.59604	-34.4451	2.32 ± 0.29	134 ± 6	141 ± 7	103 ± 5	1
354	HATLAS J232625.3-323102	351.6056	-32.51714	2.27 ± 0.29	82 ± 5	86 ± 6	61 ± 4	0
355	HATLAS J232626.7-323135	351.61138	-32.52636	2.29 ± 0.29	82 ± 5	86 ± 6	61 ± 4	0
356	HATLAS J232827.6-294831	352.11494	-29.80854	1.84 ± 0.27	122 ± 7	111 ± 6	68 ± 4	1
357	HATLAS J232900.6-321744	352.25263	-32.29559	2.98 ± 0.33	105 ± 5	127 ± 6	112 ± 5	1
358	HATLAS J233128.0-342139	352.86648	-34.36076	1.37 ± 0.26	135 ± 8	103 ± 6	53 ± 3	1
359	HATLAS J233354.6-351627	353.47754	-35.27406	1 ± 0.27	145 ± 9	93 ± 6	42 ± 3	2
360	HATLAS J233437.0-303301	353.65431	-30.55039	1.19 ± 0.26	206 ± 8	145 ± 6	70 ± 3	1
361	HATLAS J233837.3-345509	354.65525	-34.91925	1.31 ± 0.26	132 ± 7	99 ± 5	50 ± 3	-1
362	HATLAS J234357.7-351724	355.99055	-35.28988	1.98 ± 0.28	234 ± 7	224 ± 7	144 ± 4	0
363	HATLAS J234418.1-303936	356.07554	-30.66013	3.13 ± 0.34	133 ± 5	166 ± 6	152 ± 6	0
364	HATLAS J234625.9-332615	356.60793	-33.43741	1.54 ± 0.27	164 ± 9	135 ± 7	74 ± 4	0
365	HATLAS J234955.7-330833	357.48208	-33.14252	2.87 ± 0.32	89 ± 6	106 ± 7	91 ± 6	-1
366	HATLAS J235154.9-355243	357.97876	-35.87856	1.2 ± 0.26	180 ± 8	128 ± 6	62 ± 3	0
367	HATLAS J235517.9-293517	358.82443	-29.58814	1.9 ± 0.28	87 ± 6	81 ± 5	51 ± 3	-1
368	HATLAS J235623.1-354119	359.09612	-35.68855	2.99 ± 0.33	120 ± 6	145 ± 7	129 ± 6	-1
369	HATLAS J235827.7-323244	359.61534	-32.54559	3.53 ± 0.36	100 ± 5	134 ± 7	135 ± 7	-1
370	HATLAS J235910.1-325701	359.79209	-32.95035	1.57 ± 0.27	122 ± 8	101 ± 7	56 ± 4	-1

Table B.1 (cont.)

No.	H-ATLAS name	α (deg)	δ (deg)	$z_{MMF}^{unbiased}$	S_{MMF} (mJy/beam)			Reliability
					250 μm	350 μm	500 μm	
1	HATLAS J083359.4+005008	128.497828	0.835735	3.81 ± 0.37	22 ± 4	31 ± 6	33 ± 6	-1
2	HATLAS J083436.7+012907	128.653175	1.485354	4.41 ± 0.39	17 ± 3	28 ± 5	33 ± 6	-1
3	HATLAS J083700.1+013111	129.250571	1.519785	6.03 ± 0.34	14 ± 2	39 ± 5	60 ± 7	-1
4	HATLAS J084116.1+002619	130.317356	0.438641	5.01 ± 0.4	15 ± 3	29 ± 5	37 ± 7	-1
5	HATLAS J084213.4-000531	130.556008	-0.091959	3.76 ± 0.37	26 ± 5	36 ± 6	38 ± 7	-1
6	HATLAS J084644.6+002352	131.686002	0.397871	3.96 ± 0.38	18 ± 4	26 ± 5	28 ± 6	1
7	HATLAS J085340.9+024044	133.420602	2.679046	5.14 ± 0.39	13 ± 2	26 ± 5	35 ± 7	-1
8	HATLAS J085534.2+005258	133.892785	0.882992	5.29 ± 0.39	12 ± 2	26 ± 5	35 ± 7	-1
9	HATLAS J090029.8+001653	135.12445	0.281402	5.17 ± 0.39	11 ± 2	22 ± 4	29 ± 5	-1
10	HATLAS J090045.5+004131	135.189985	0.69217	6.35 ± 0.3	8 ± 1	25 ± 4	40 ± 6	0
11	HATLAS J090304.5-004616	135.769163	-0.771248	6.35 ± 0.3	7 ± 1	23 ± 4	37 ± 6	-1
12	HATLAS J090329.6+000753	135.873578	0.131548	4.89 ± 0.4	18 ± 3	33 ± 5	43 ± 6	-1
13	HATLAS J090540.0+020347	136.417044	2.063127	3.65 ± 0.36	32 ± 5	44 ± 7	45 ± 7	-1
14	HATLAS J090908.6-010140	137.286131	-1.027792	4.33 ± 0.39	20 ± 4	31 ± 6	36 ± 7	-1
15	HATLAS J091019.4+011755	137.581142	1.298769	4.04 ± 0.38	16 ± 3	24 ± 4	26 ± 5	-1
16	HATLAS J091025.9+002116	137.60821	0.354684	4.34 ± 0.39	16 ± 3	25 ± 5	29 ± 5	-1
17	HATLAS J091617.1+010036	139.071573	1.010031	4.97 ± 0.4	14 ± 3	26 ± 5	34 ± 6	0
18	HATLAS J091733.4+005054	139.389273	0.848382	4.34 ± 0.39	21 ± 3	33 ± 5	39 ± 6	-1
19	HATLAS J091902.6-003905	139.760959	-0.651409	5.8 ± 0.36	8 ± 2	21 ± 4	31 ± 6	-1
20	HATLAS J092215.4+002920	140.564326	0.489096	3.93 ± 0.38	22 ± 4	32 ± 6	35 ± 7	0
21	HATLAS J092406.5+003901	141.02734	0.65037	4.2 ± 0.39	17 ± 3	26 ± 5	29 ± 6	-1
22	HATLAS J092447.2+005854	141.19684	0.98167	3.61 ± 0.36	23 ± 4	32 ± 5	32 ± 6	1
23	HATLAS J092554.4+002846	141.47684	0.47952	5.12 ± 0.39	14 ± 2	27 ± 4	36 ± 5	-1
24	HATLAS J113009.3+004659	172.53891	0.78328	5.29 ± 0.39	10 ± 2	22 ± 4	30 ± 5	-1
25	HATLAS J113335.1+002616	173.39658	0.43789	4.95 ± 0.4	15 ± 3	28 ± 5	36 ± 7	-1
26	HATLAS J113650.3+010128	174.20998	1.0247	4.75 ± 0.4	17 ± 3	30 ± 5	37 ± 6	0
27	HATLAS J113942.6+001742	174.92773	0.29517	5.29 ± 0.39	13 ± 2	28 ± 5	38 ± 7	0
28	HATLAS J114038.8-022804	175.16169	-2.46805	5.38 ± 0.38	13 ± 2	27 ± 5	38 ± 7	-1
29	HATLAS J114137.2-001154	175.40502	-0.19846	4.68 ± 0.4	17 ± 3	30 ± 5	37 ± 7	-1
30	HATLAS J114257.7-005624	175.74048	-0.94008	4.3 ± 0.39	20 ± 3	31 ± 5	37 ± 6	0
31	HATLAS J114350.3-005210	175.95963	-0.86963	4.75 ± 0.4	14 ± 3	26 ± 5	32 ± 6	-1
32	HATLAS J114353.5+001250	175.97311	0.21399	5.39 ± 0.38	11 ± 2	24 ± 4	33 ± 6	1
33	HATLAS J114412.1+001812	176.05062	0.30345	4.67 ± 0.4	17 ± 3	30 ± 5	37 ± 6	-1
34	HATLAS J114420.2-020218	176.08449	-2.0386	4.93 ± 0.4	13 ± 3	25 ± 5	33 ± 6	0
35	HATLAS J114724.8-012743	176.85339	-1.46213	5.48 ± 0.38	10 ± 2	23 ± 4	32 ± 6	-1
36	HATLAS J115306.5-014248	178.2772	-1.71345	4.84 ± 0.4	16 ± 3	30 ± 6	38 ± 7	-1
37	HATLAS J115319.3-022936	178.3305	-2.49336	5.43 ± 0.38	12 ± 2	27 ± 5	38 ± 7	-1
38	HATLAS J115435.6-013057	178.64854	-1.51593	3.72 ± 0.37	25 ± 5	34 ± 6	36 ± 7	-1
39	HATLAS J115454.5-010322	178.72738	-1.05624	3.85 ± 0.37	25 ± 4	36 ± 5	39 ± 6	2
40	HATLAS J115606.9+012458	179.02879	1.41632	4.16 ± 0.39	17 ± 3	25 ± 5	29 ± 6	-1
41	HATLAS J115614.0+013900	179.0586	1.65024	5.89 ± 0.35	11 ± 2	30 ± 5	44 ± 7	-1
42	HATLAS J115633.8+014521	179.14101	1.75603	3.98 ± 0.38	15 ± 3	22 ± 4	24 ± 5	-1
43	HATLAS J115719.3-010052	179.33047	-1.0146	3.89 ± 0.38	21 ± 4	30 ± 6	33 ± 6	1
44	HATLAS J115803.2-004836	179.51362	-0.81008	3.84 ± 0.37	22 ± 4	32 ± 6	34 ± 6	-1
45	HATLAS J120022.8-004025	180.09541	-0.67377	3.91 ± 0.38	18 ± 4	26 ± 5	28 ± 6	-1
46	HATLAS J120904.8-010524	182.27007	-1.09008	5.51 ± 0.38	14 ± 2	33 ± 5	47 ± 7	-1
47	HATLAS J121004.6+011057	182.51954	1.18271	5.95 ± 0.34	8 ± 2	22 ± 4	34 ± 6	-1
48	HATLAS J121023.4+002621	182.5978	0.43918	6.35 ± 0.3	7 ± 1	21 ± 3	34 ± 6	0
49	HATLAS J121638.3-005132	184.15999	-0.85916	3.89 ± 0.38	20 ± 4	28 ± 5	31 ± 6	1
50	HATLAS J121646.7-000233	184.19499	-0.04274	5.59 ± 0.37	11 ± 2	27 ± 5	38 ± 7	0
51	HATLAS J121656.1-012557	184.23383	-1.43273	4.71 ± 0.4	17 ± 3	30 ± 5	37 ± 6	-1
52	HATLAS J121740.3-021133	184.41806	-2.19264	3.85 ± 0.37	20 ± 4	28 ± 5	30 ± 6	-1
53	HATLAS J121757.1-010815	184.48827	-1.13767	4.38 ± 0.39	20 ± 4	32 ± 6	38 ± 7	0
54	HATLAS J121803.7+005743	184.5156	0.96205	3.55 ± 0.36	23 ± 4	31 ± 5	31 ± 6	-1
55	HATLAS J122113.8-003149	185.30778	-0.53044	5.03 ± 0.39	17 ± 3	32 ± 6	42 ± 7	-1
56	HATLAS J122130.1-013308	185.37583	-1.5524	3.85 ± 0.37	22 ± 4	31 ± 6	33 ± 6	-1
57	HATLAS J122244.7+002320	185.68666	0.38914	4.27 ± 0.39	23 ± 4	36 ± 6	42 ± 7	-1
58	HATLAS J140926.1-012136	212.35893	-1.36003	4.21 ± 0.39	19 ± 4	29 ± 6	34 ± 6	0
59	HATLAS J140938.5+012832	212.4106	1.47561	4 ± 0.38	20 ± 4	29 ± 6	33 ± 6	-1
60	HATLAS J140959.8-015337	212.49924	-1.89362	4.19 ± 0.39	18 ± 3	27 ± 5	31 ± 6	-1
61	HATLAS J141143.1-001251	212.92996	-0.21442	4.02 ± 0.38	19 ± 4	28 ± 5	32 ± 6	0
62	HATLAS J141416.4+023453	213.56857	2.5816	4.37 ± 0.39	20 ± 3	32 ± 5	38 ± 6	-1
63	HATLAS J142141.4+011904	215.42273	1.31798	5.77 ± 0.36	10 ± 2	25 ± 5	36 ± 7	-1
64	HATLAS J142516.7-010213	216.31993	-1.0371	4.03 ± 0.38	20 ± 4	30 ± 6	33 ± 7	-1
65	HATLAS J142710.6+013806	216.79439	1.63525	3.79 ± 0.37	28 ± 4	40 ± 6	42 ± 6	1

Table B.2: 237 H-ATLAS sources that form our 500 μm-riser sample.

No.	H-ATLAS name	α (deg)	δ (deg)	$z_{MMF}^{unbiased}$	S_{MMF} (mJy/beam)			Reliability
					250 μm	350 μm	500 μm	
66	HATLAS J142734.7+000835	216.8947	0.14321	4.92 ± 0.4	16 ± 3	30 ± 5	38 ± 7	-1
67	HATLAS J142756.2+000915	216.98433	0.15432	3.95 ± 0.38	21 ± 4	31 ± 5	34 ± 6	-1
68	HATLAS J143447.5-011809	218.69818	-1.30268	4.11 ± 0.38	27 ± 4	40 ± 6	45 ± 7	1
69	HATLAS J143513.7-011042	218.80728	-1.17844	5.22 ± 0.39	13 ± 2	26 ± 5	35 ± 6	-1
70	HATLAS J143727.4-013459	219.36453	-1.58314	3.83 ± 0.37	21 ± 4	30 ± 6	32 ± 6	-1
71	HATLAS J143919.5+015908	219.83143	1.98582	5.3 ± 0.39	12 ± 2	26 ± 5	35 ± 7	-1
72	HATLAS J143953.7+012507	219.97409	1.41863	5.17 ± 0.39	12 ± 2	25 ± 4	34 ± 6	1
73	HATLAS J144128.0+001355	220.36674	0.23211	5.45 ± 0.38	9 ± 2	20 ± 4	28 ± 5	-1
74	HATLAS J144410.5+004845	221.04401	0.81252	4.2 ± 0.39	17 ± 3	26 ± 5	30 ± 6	-1
75	HATLAS J144753.9+015400	221.97496	1.90014	3.94 ± 0.38	21 ± 4	31 ± 5	34 ± 6	1
76	HATLAS J144856.6+024806	222.23606	2.80174	4.21 ± 0.39	22 ± 3	34 ± 5	39 ± 6	-1
77	HATLAS J145035.8+000611	222.64936	0.10308	5.07 ± 0.39	14 ± 3	27 ± 5	35 ± 6	-1
78	HATLAS J145039.4-003925	222.66433	-0.65698	3.63 ± 0.36	23 ± 4	31 ± 6	32 ± 6	0
79	HATLAS J000006.6-303443	0.02755	-30.57851	4.72 ± 0.4	15 ± 3	26 ± 5	32 ± 6	0
80	HATLAS J000308.6-331749	0.7857	-33.2969	4.37 ± 0.39	18 ± 3	28 ± 5	33 ± 6	-1
81	HATLAS J000442.1-324008	1.1754	-32.66889	3.98 ± 0.38	33 ± 4	48 ± 6	53 ± 6	1
82	HATLAS J000824.5-340548	2.10221	-34.09667	5.2 ± 0.39	13 ± 2	27 ± 5	37 ± 6	-1
83	HATLAS J001305.3-304540	3.27212	-30.76105	3.84 ± 0.37	22 ± 4	31 ± 5	33 ± 6	1
84	HATLAS J001534.5-304821	3.89389	-30.80579	4.56 ± 0.4	18 ± 3	30 ± 5	37 ± 6	-1
85	HATLAS J001748.4-335213	4.45155	-33.87021	3.88 ± 0.37	20 ± 4	28 ± 5	31 ± 6	-1
86	HATLAS J001917.8-333739	4.82428	-33.62757	4.77 ± 0.4	15 ± 3	27 ± 5	34 ± 6	-1
87	HATLAS J002028.3-353119	5.11775	-35.52207	4.99 ± 0.4	16 ± 3	30 ± 6	39 ± 7	2
88	HATLAS J002237.1-300722	5.65461	-30.12283	5.27 ± 0.39	12 ± 2	26 ± 5	35 ± 7	1
89	HATLAS J002300.8-333258	5.75341	-33.54936	3.89 ± 0.38	22 ± 4	31 ± 6	34 ± 6	-1
90	HATLAS J002347.4-322513	5.94732	-32.42015	4.2 ± 0.39	20 ± 3	30 ± 5	34 ± 6	-1
91	HATLAS J002435.7-295057	6.14872	-29.84914	4.09 ± 0.38	22 ± 4	32 ± 6	37 ± 7	-1
92	HATLAS J002546.6-300507	6.444	-30.08534	4.44 ± 0.39	17 ± 3	27 ± 5	32 ± 6	-1
93	HATLAS J002603.6-305948	6.51518	-30.99678	4.52 ± 0.39	17 ± 3	28 ± 5	34 ± 6	-1
94	HATLAS J003152.9-284343	7.97048	-28.72848	4.5 ± 0.39	17 ± 3	29 ± 6	34 ± 7	-1
95	HATLAS J003214.7-295656	8.06122	-29.9489	4.95 ± 0.4	15 ± 3	29 ± 5	38 ± 7	-1
96	HATLAS J003247.6-313214	8.19844	-31.53713	3.96 ± 0.38	18 ± 4	26 ± 5	29 ± 6	-1
97	HATLAS J003423.2-301146	8.59675	-30.196	5.16 ± 0.39	15 ± 2	30 ± 5	39 ± 6	-1
98	HATLAS J003431.0-275557	8.62929	-27.93253	3.69 ± 0.37	25 ± 5	35 ± 6	36 ± 7	-1
99	HATLAS J003458.8-324439	8.74482	-32.74422	4.29 ± 0.39	21 ± 4	33 ± 6	38 ± 6	0
100	HATLAS J003506.6-280621	8.77764	-28.10575	5.32 ± 0.39	12 ± 2	26 ± 5	35 ± 7	-1
101	HATLAS J003603.4-331114	9.01405	-33.18734	4.38 ± 0.39	18 ± 3	29 ± 5	35 ± 6	-1
102	HATLAS J003746.4-304516	9.44345	-30.75435	3.95 ± 0.38	25 ± 4	36 ± 6	40 ± 7	2
103	HATLAS J003830.8-293604	9.62824	-29.60115	4.69 ± 0.4	16 ± 3	29 ± 5	35 ± 6	-1
104	HATLAS J004135.5-301743	10.39791	-30.29526	3.91 ± 0.38	21 ± 4	31 ± 5	33 ± 6	-1
105	HATLAS J004202.4-275650	10.51004	-27.9471	6.34 ± 0.31	8 ± 1	25 ± 5	41 ± 7	-1
106	HATLAS J004218.4-302835	10.57666	-30.47626	4.9 ± 0.4	13 ± 2	25 ± 4	32 ± 5	0
107	HATLAS J004226.7-295508	10.61104	-29.91892	4.3 ± 0.39	19 ± 3	31 ± 5	36 ± 6	-1
108	HATLAS J004336.8-294629	10.90347	-29.77478	4.84 ± 0.4	14 ± 3	25 ± 5	32 ± 6	-1
109	HATLAS J004531.3-273655	11.38037	-27.61537	4.65 ± 0.4	17 ± 3	29 ± 5	35 ± 6	-1
110	HATLAS J004615.0-321825	11.56239	-32.30682	5.65 ± 0.37	14 ± 2	33 ± 5	48 ± 7	-1
111	HATLAS J004920.9-310555	12.33711	-31.09873	4.96 ± 0.4	17 ± 3	32 ± 5	41 ± 7	-1
112	HATLAS J005133.0-285422	12.88762	-28.90616	3.91 ± 0.38	24 ± 4	35 ± 6	38 ± 7	-1
113	HATLAS J005216.5-283325	13.06888	-28.55688	4.73 ± 0.4	16 ± 3	29 ± 6	36 ± 7	-1
114	HATLAS J005906.3-284622	14.77608	-28.77279	3.79 ± 0.37	24 ± 4	34 ± 6	36 ± 6	-1
115	HATLAS J005940.0-312519	14.9166	-31.42208	3.59 ± 0.36	21 ± 4	29 ± 5	29 ± 5	-1
116	HATLAS J010105.2-333447	15.27156	-33.57965	4.96 ± 0.4	16 ± 3	30 ± 5	38 ± 6	-1
117	HATLAS J010126.1-313311	15.35888	-31.55307	5.11 ± 0.39	14 ± 3	29 ± 5	38 ± 7	-1
118	HATLAS J010137.7-310438	15.40711	-31.07724	4.37 ± 0.39	21 ± 4	34 ± 6	40 ± 7	2
119	HATLAS J010216.1-331339	15.56692	-33.22763	6.35 ± 0.3	7 ± 1	23 ± 4	38 ± 7	-1
120	HATLAS J010322.9-290946	15.84526	-29.16287	6.35 ± 0.3	7 ± 1	23 ± 4	38 ± 7	-1
121	HATLAS J010641.5-295536	16.67308	-29.92658	4.34 ± 0.39	21 ± 4	33 ± 6	39 ± 7	-1
122	HATLAS J011217.8-325839	18.0741	-32.97742	4.97 ± 0.4	16 ± 3	31 ± 6	40 ± 7	-1
123	HATLAS J011459.2-295600	18.74664	-29.93332	4.19 ± 0.39	19 ± 3	29 ± 5	33 ± 6	0
124	HATLAS J011502.9-311639	18.76194	-31.27742	3.55 ± 0.36	23 ± 4	31 ± 6	32 ± 6	-1
125	HATLAS J011717.8-313434	19.3242	-31.57618	5.63 ± 0.37	12 ± 2	29 ± 5	41 ± 7	0
126	HATLAS J011756.9-312505	19.48717	-31.418	5.03 ± 0.39	14 ± 3	26 ± 5	34 ± 7	-1
127	HATLAS J011931.6-275452	19.88175	-27.91452	4.49 ± 0.39	16 ± 3	26 ± 5	31 ± 6	1
128	HATLAS J012048.8-305828	20.20345	-30.97436	5.49 ± 0.38	13 ± 2	30 ± 6	42 ± 8	-1
129	HATLAS J012050.3-331932	20.2097	-33.32547	5.58 ± 0.37	11 ± 2	25 ± 4	36 ± 6	-1
130	HATLAS J012213.2-302146	20.55511	-30.36281	4.08 ± 0.38	18 ± 3	27 ± 5	30 ± 5	0
131	HATLAS J012410.4-274606	21.04342	-27.76847	4.42 ± 0.39	26 ± 3	42 ± 6	50 ± 7	0

Table B.2 (cont.)

No.	H-ATLAS name	α (deg)	δ (deg)	$z_{MMF}^{unbiased}$	S_{MMF} (mJy/beam)			Reliability
					250 μm	350 μm	500 μm	
132	HATLAS J012435.8-311231	21.14933	-31.20873	5.33 ± 0.39	15 ± 2	33 ± 5	45 ± 7	-1
133	HATLAS J012919.5-334807	22.33135	-33.80198	5.8 ± 0.36	10 ± 2	25 ± 5	36 ± 7	0
134	HATLAS J013155.8-311145	22.98238	-31.19597	5.26 ± 0.39	11 ± 2	22 ± 4	30 ± 6	0
135	HATLAS J013437.1-294945	23.65446	-29.82918	4.14 ± 0.39	20 ± 4	30 ± 6	34 ± 6	-1
136	HATLAS J013600.1-331414	24.00026	-33.23716	4.64 ± 0.4	19 ± 3	32 ± 5	40 ± 6	1
137	HATLAS J124559.5+325855	191.49782	32.98193	5.04 ± 0.39	14 ± 3	26 ± 5	34 ± 7	-1
138	HATLAS J124818.4+331525	192.07671	33.25682	6.22 ± 0.32	8 ± 2	24 ± 5	38 ± 8	-1
139	HATLAS J124821.7+324712	192.09048	32.78667	3.74 ± 0.37	21 ± 4	29 ± 6	31 ± 6	-1
140	HATLAS J125005.3+295627	192.52188	29.9409	4.94 ± 0.4	13 ± 3	25 ± 5	33 ± 6	-1
141	HATLAS J125041.9+323216	192.67448	32.53773	4.54 ± 0.4	17 ± 3	29 ± 6	35 ± 7	1
142	HATLAS J125048.4+263632	192.70163	26.60883	3.6 ± 0.36	22 ± 4	30 ± 5	31 ± 6	-1
143	HATLAS J125335.7+231935	193.39875	23.32631	3.59 ± 0.36	31 ± 5	42 ± 6	43 ± 6	1
144	HATLAS J125413.6+221132	193.55682	22.19217	4.11 ± 0.38	24 ± 3	35 ± 5	40 ± 6	-1
145	HATLAS J125547.7+224229	193.9489	22.70793	5.64 ± 0.37	11 ± 2	26 ± 5	38 ± 7	-1
146	HATLAS J125729.5+291649	194.37289	29.28026	5.48 ± 0.38	13 ± 2	30 ± 5	42 ± 7	1
147	HATLAS J125858.8+224602	194.74495	22.76724	6.35 ± 0.3	7 ± 1	22 ± 4	36 ± 6	-1
148	HATLAS J130233.7+264025	195.64042	26.67359	3.85 ± 0.37	31 ± 4	44 ± 6	47 ± 6	-1
149	HATLAS J130303.9+225541	195.76613	22.92814	4.74 ± 0.4	17 ± 3	30 ± 5	37 ± 6	0
150	HATLAS J130317.0+305015	195.82088	30.8375	4.57 ± 0.4	17 ± 3	29 ± 5	36 ± 6	1
151	HATLAS J130337.1+311655	195.90467	31.28193	5.34 ± 0.39	11 ± 2	24 ± 5	33 ± 6	1
152	HATLAS J130406.1+243843	196.02545	24.64537	4.79 ± 0.4	15 ± 3	27 ± 5	33 ± 6	0
153	HATLAS J130548.2+240824	196.45068	24.14001	4.52 ± 0.39	16 ± 3	26 ± 5	31 ± 6	1
154	HATLAS J130615.7+263834	196.56545	26.6429	3.86 ± 0.37	20 ± 4	28 ± 5	30 ± 6	1
155	HATLAS J130835.6+243036	197.14837	24.50999	5.12 ± 0.39	16 ± 3	32 ± 6	42 ± 8	-1
156	HATLAS J131051.0+243309	197.71234	24.55262	3.56 ± 0.36	26 ± 5	35 ± 6	35 ± 6	-1
157	HATLAS J131133.9+273020	197.89108	27.50556	4.08 ± 0.38	20 ± 4	30 ± 6	34 ± 6	-1
158	HATLAS J131218.0+330614	198.075	33.10383	4.49 ± 0.39	18 ± 3	29 ± 5	35 ± 6	-1
159	HATLAS J131354.8+285257	198.47817	28.8826	4.44 ± 0.39	17 ± 3	27 ± 5	33 ± 6	-1
160	HATLAS J131457.5+334250	198.73942	33.71387	5.23 ± 0.39	12 ± 2	26 ± 5	34 ± 7	-1
161	HATLAS J131650.8+254045	199.2116	25.6793	5.18 ± 0.39	13 ± 2	26 ± 5	35 ± 7	0
162	HATLAS J131732.7+225516	199.38609	22.9211	3.83 ± 0.37	25 ± 5	35 ± 7	37 ± 7	-1
163	HATLAS J131800.5+350909	199.50228	35.15254	4.27 ± 0.39	21 ± 4	32 ± 6	38 ± 7	1
164	HATLAS J131818.7+261439	199.578	26.2442	6.01 ± 0.34	11 ± 1	30 ± 4	46 ± 6	0
165	HATLAS J131953.4+231355	199.9723	23.23182	3.7 ± 0.37	24 ± 4	33 ± 5	34 ± 6	-1
166	HATLAS J132037.2+272629	200.15492	27.44151	3.68 ± 0.37	21 ± 4	29 ± 6	30 ± 6	-1
167	HATLAS J132122.8+261443	200.34519	26.24517	3.8 ± 0.37	28 ± 4	40 ± 6	42 ± 7	-1
168	HATLAS J132158.0+251200	200.49181	25.19987	3.93 ± 0.38	23 ± 4	33 ± 6	36 ± 7	0
169	HATLAS J132240.6+314953	200.66932	31.83131	3.55 ± 0.36	20 ± 4	27 ± 5	28 ± 5	-1
170	HATLAS J132247.9+234915	200.6994	23.8208	4.66 ± 0.4	14 ± 3	25 ± 5	31 ± 6	-1
171	HATLAS J132306.3+244457	200.77612	24.74909	4.23 ± 0.39	23 ± 4	36 ± 5	41 ± 6	-1
172	HATLAS J132354.5+345052	200.97727	34.84778	4.21 ± 0.39	19 ± 4	30 ± 5	34 ± 6	-1
173	HATLAS J132442.3+262451	201.17616	26.41423	4.79 ± 0.4	19 ± 3	35 ± 5	43 ± 7	-1
174	HATLAS J132525.3+310641	201.35521	31.11126	5.44 ± 0.38	12 ± 2	27 ± 5	38 ± 7	-1
175	HATLAS J132810.7+253419	202.04466	25.57205	4.13 ± 0.38	22 ± 4	33 ± 6	37 ± 6	1
176	HATLAS J132854.8+243545	202.2282	24.59587	4.35 ± 0.39	17 ± 3	28 ± 5	32 ± 6	-1
177	HATLAS J132928.9+274016	202.37052	27.67114	4.45 ± 0.39	17 ± 3	28 ± 5	33 ± 6	-1
178	HATLAS J132932.8+243125	202.38677	24.52368	3.59 ± 0.36	23 ± 4	31 ± 6	32 ± 6	-1
179	HATLAS J132949.9+283751	202.45776	28.63093	4.09 ± 0.38	22 ± 4	32 ± 6	36 ± 6	-1
180	HATLAS J133102.3+334218	202.75967	33.70496	4.33 ± 0.39	17 ± 3	27 ± 5	32 ± 6	-1
181	HATLAS J133115.7+322657	202.81525	32.44905	4.87 ± 0.4	14 ± 3	26 ± 5	32 ± 6	-1
182	HATLAS J133131.1+272503	202.87978	27.41749	5.02 ± 0.39	17 ± 3	33 ± 5	43 ± 6	-1
183	HATLAS J133245.4+260531	203.18921	26.09199	4.2 ± 0.39	20 ± 4	30 ± 5	34 ± 6	-1
184	HATLAS J133404.6+321017	203.51898	32.17136	4.57 ± 0.4	16 ± 3	27 ± 5	33 ± 6	-1
185	HATLAS J133406.6+300930	203.5275	30.15842	4.9 ± 0.4	13 ± 2	24 ± 4	31 ± 6	0
186	HATLAS J133416.2+284601	203.56736	28.76707	4.24 ± 0.39	17 ± 3	26 ± 5	30 ± 6	-1
187	HATLAS J133437.9+244547	203.65811	24.76305	4.33 ± 0.39	19 ± 4	30 ± 6	35 ± 7	-1
188	HATLAS J133447.6+302524	203.69846	30.42342	3.92 ± 0.38	21 ± 4	31 ± 5	34 ± 6	-1
189	HATLAS J133734.5+233433	204.39369	23.57589	5.06 ± 0.39	14 ± 3	28 ± 5	37 ± 7	-1
190	HATLAS J133803.8+333002	204.51599	33.50048	5.12 ± 0.39	13 ± 2	26 ± 5	35 ± 7	1
191	HATLAS J133812.8+321941	204.55317	32.32804	3.61 ± 0.36	21 ± 4	29 ± 6	30 ± 6	-1
192	HATLAS J134207.7+352858	205.53188	35.48283	4.08 ± 0.38	22 ± 4	33 ± 5	37 ± 6	-1
193	HATLAS J134256.5+255754	205.73562	25.96504	3.61 ± 0.36	23 ± 4	31 ± 6	32 ± 6	-1
194	HATLAS J134301.8+232138	205.75738	23.36044	5.76 ± 0.36	10 ± 2	24 ± 5	36 ± 7	-1
195	HATLAS J134422.2+265845	206.09236	26.97918	4.48 ± 0.39	19 ± 3	31 ± 6	37 ± 7	-1
196	HATLAS J134515.4+304455	206.31416	30.74855	3.81 ± 0.37	22 ± 4	31 ± 6	33 ± 6	2

Table B.2 (cont.)

No.	H-ATLAS name	α (deg)	δ (deg)	$z_{MMF}^{unbiased}$	S_{MMF} (mJy/beam)			Reliability
					250 μm	350 μm	500 μm	
197	HATLAS J134531.6+281303	206.38154	28.21738	5.06 ± 0.39	13 ± 2	25 ± 5	33 ± 6	-1
198	HATLAS J134534.3+284050	206.39279	28.68045	5.6 ± 0.37	12 ± 2	28 ± 5	40 ± 7	-1
199	HATLAS J134622.9+233256	206.5955	23.54898	5.08 ± 0.39	16 ± 3	32 ± 5	42 ± 7	-1
200	HATLAS J134735.5+245809	206.89795	24.9692	4.94 ± 0.4	16 ± 3	30 ± 6	39 ± 7	-1
201	HATLAS J134944.5+273609	207.43532	27.60259	3.87 ± 0.37	22 ± 4	31 ± 6	34 ± 6	-1
202	HATLAS J222645.4-292709	336.68912	-29.45246	4.46 ± 0.39	23 ± 4	37 ± 6	44 ± 8	-1
203	HATLAS J224121.3-305125	340.33885	-30.85684	4.17 ± 0.39	18 ± 4	28 ± 5	32 ± 6	-1
204	HATLAS J224211.1-335351	340.5464	-33.89737	5.03 ± 0.39	17 ± 3	34 ± 5	44 ± 7	-1
205	HATLAS J224221.7-350505	340.59061	-35.08485	3.72 ± 0.37	22 ± 4	31 ± 6	33 ± 6	-1
206	HATLAS J224625.4-352050	341.60585	-35.34713	3.54 ± 0.36	23 ± 4	31 ± 5	31 ± 5	-1
207	HATLAS J225017.9-321548	342.57478	-32.26341	3.97 ± 0.38	25 ± 4	36 ± 5	39 ± 6	-1
208	HATLAS J225416.8-293024	343.57008	-29.50654	5.76 ± 0.36	10 ± 2	24 ± 4	36 ± 6	-1
209	HATLAS J225929.3-292611	344.87215	-29.43648	5.3 ± 0.39	13 ± 2	27 ± 5	37 ± 6	-1
210	HATLAS J230233.6-324821	345.63999	-32.80592	6.35 ± 0.3	7 ± 1	22 ± 4	35 ± 7	-1
211	HATLAS J230421.2-340806	346.08818	-34.13486	4.24 ± 0.39	19 ± 3	30 ± 5	34 ± 6	-1
212	HATLAS J230604.4-324951	346.51814	-32.83094	4.74 ± 0.4	14 ± 3	25 ± 5	31 ± 6	-1
213	HATLAS J230648.5-295046	346.70199	-29.8462	4.89 ± 0.4	15 ± 3	28 ± 5	35 ± 6	-1
214	HATLAS J230718.6-325452	346.82762	-32.91445	4.98 ± 0.4	15 ± 3	29 ± 6	38 ± 7	0
215	HATLAS J231057.8-334828	347.74092	-33.80789	4.73 ± 0.4	15 ± 3	26 ± 5	32 ± 6	0
216	HATLAS J232149.3-351916	350.45546	-35.32113	5.14 ± 0.39	13 ± 2	27 ± 5	36 ± 7	-1
217	HATLAS J232322.3-355631	350.84293	-35.942	3.8 ± 0.37	26 ± 4	36 ± 6	39 ± 6	-1
218	HATLAS J232505.8-331012	351.2741	-33.16989	5.7 ± 0.37	8 ± 2	20 ± 4	30 ± 6	-1
219	HATLAS J232550.2-320153	351.45898	-32.03141	3.82 ± 0.37	21 ± 4	30 ± 5	32 ± 6	2
220	HATLAS J232631.8-340353	351.63244	-34.06465	3.51 ± 0.36	24 ± 4	31 ± 6	32 ± 6	-1
221	HATLAS J232954.3-300820	352.47639	-30.13893	3.62 ± 0.36	22 ± 4	31 ± 6	31 ± 6	0
222	HATLAS J233136.8-330350	352.90326	-33.06386	5.06 ± 0.39	14 ± 3	28 ± 5	36 ± 7	-1
223	HATLAS J233245.1-354618	353.18775	-35.77158	5.33 ± 0.39	15 ± 3	33 ± 6	45 ± 9	-1
224	HATLAS J233254.3-345553	353.22623	-34.9313	4.52 ± 0.39	21 ± 4	34 ± 6	41 ± 7	0
225	HATLAS J233344.1-323041	353.43366	-32.51129	4.54 ± 0.4	14 ± 3	24 ± 5	29 ± 6	-1
226	HATLAS J233806.0-321923	354.52505	-32.32305	5.32 ± 0.39	12 ± 2	25 ± 5	34 ± 6	-1
227	HATLAS J233900.7-352059	354.75279	-35.34978	5.12 ± 0.39	15 ± 3	30 ± 5	39 ± 7	-1
228	HATLAS J234151.2-353451	355.46321	-35.58076	4.69 ± 0.4	16 ± 3	28 ± 6	35 ± 7	-1
229	HATLAS J234611.7-330157	356.54882	-33.03263	4.15 ± 0.39	20 ± 4	30 ± 6	34 ± 7	0
230	HATLAS J235016.4-304754	357.56828	-30.79842	4.83 ± 0.4	15 ± 3	28 ± 5	35 ± 6	1
231	HATLAS J235122.0-352516	357.84158	-35.42107	4.94 ± 0.4	14 ± 3	26 ± 5	34 ± 7	-1
232	HATLAS J235423.2-322301	358.5965	-32.38369	3.58 ± 0.36	25 ± 5	33 ± 6	34 ± 6	-1
233	HATLAS J235542.5-341236	358.92699	-34.2099	5.31 ± 0.39	11 ± 2	23 ± 4	32 ± 6	0
234	HATLAS J235608.0-351731	359.03331	-35.29186	4.39 ± 0.39	17 ± 3	27 ± 5	32 ± 5	1
235	HATLAS J235636.6-305628	359.15239	-30.94124	4.63 ± 0.4	14 ± 3	24 ± 5	30 ± 6	-1
236	HATLAS J235730.8-332701	359.3782	-33.45037	5.32 ± 0.39	12 ± 2	25 ± 5	34 ± 6	-1
237	HATLAS J235917.5-320511	359.82304	-32.08649	4.2 ± 0.39	19 ± 4	28 ± 5	33 ± 6	1

Table B.2 (cont.)

Appendix C: Candidate lensed galaxies

We show here the images of the 40 lensed galaxies candidates selected after cross-matching the robust high-z subsample of 283 high-z candidates with a sample of 1,776,242 known-low redshift galaxies from SDSS DR14. H-ATLAS sources and SDSS associations are marked with a pink and a green square, respectively. Einstein rings according to the Einstein radii estimated in Table 2.4 are drawn as red circles. Those examples predicted as lenses by one of the models derived with our CNN have their ID names highlighted in orange.

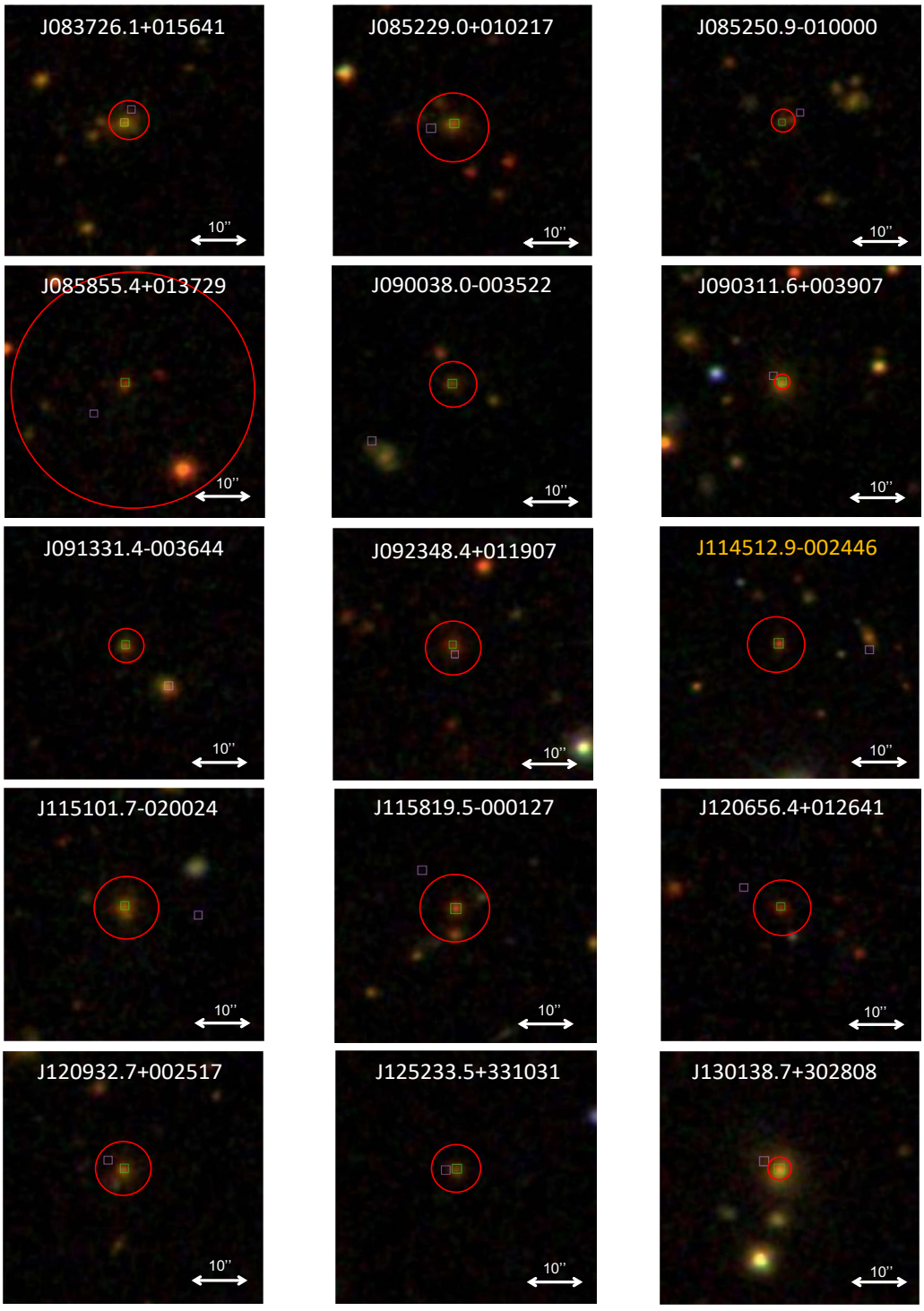


Figure C: Images (47 x 47 arcsec^2) of the 40 associations found between the robust high-z subsample of 283 high-z candidates and a sample of 1,776,242 known-low-redshift galaxies from SDSS DR14.

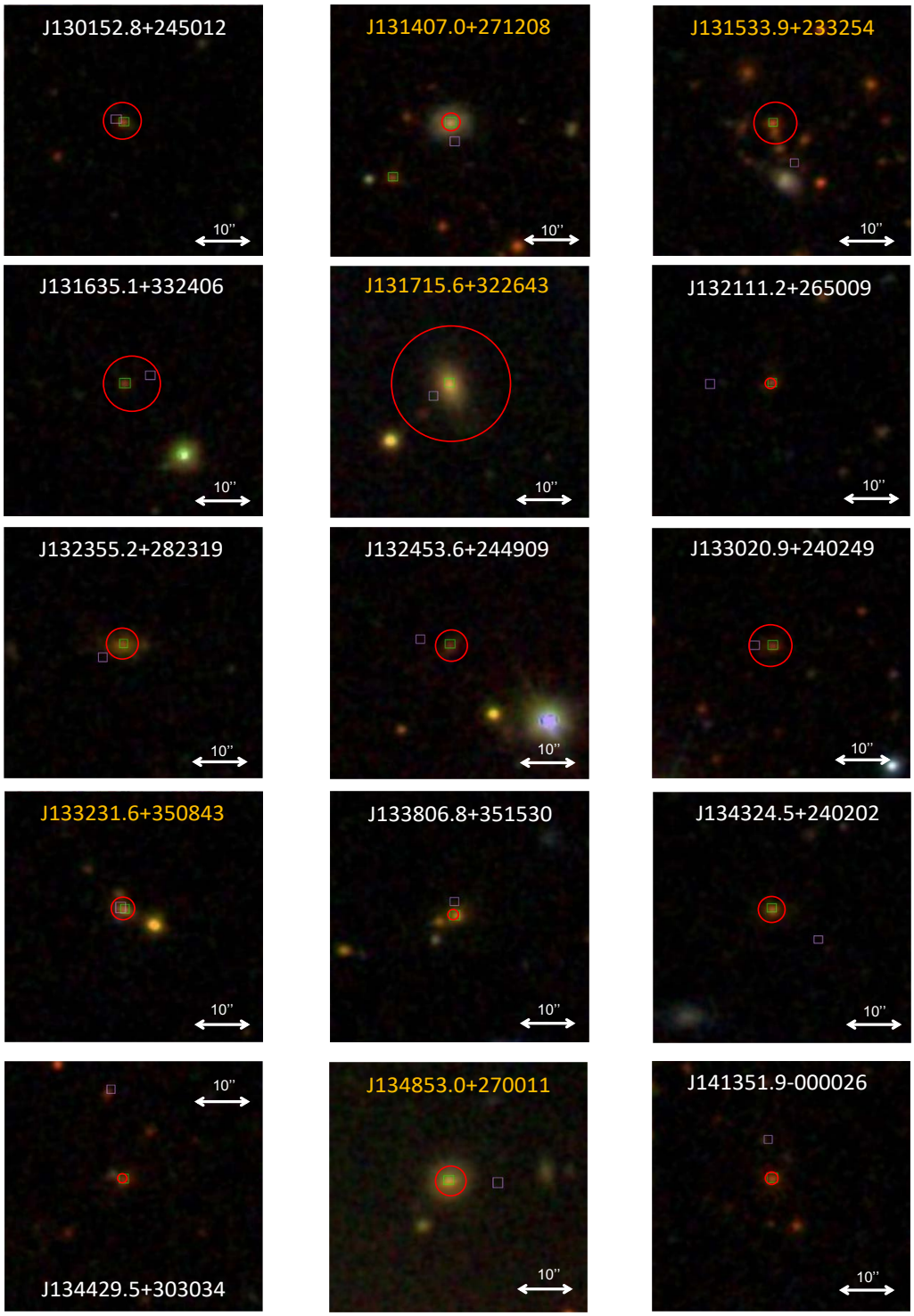


Figure C (cont.)

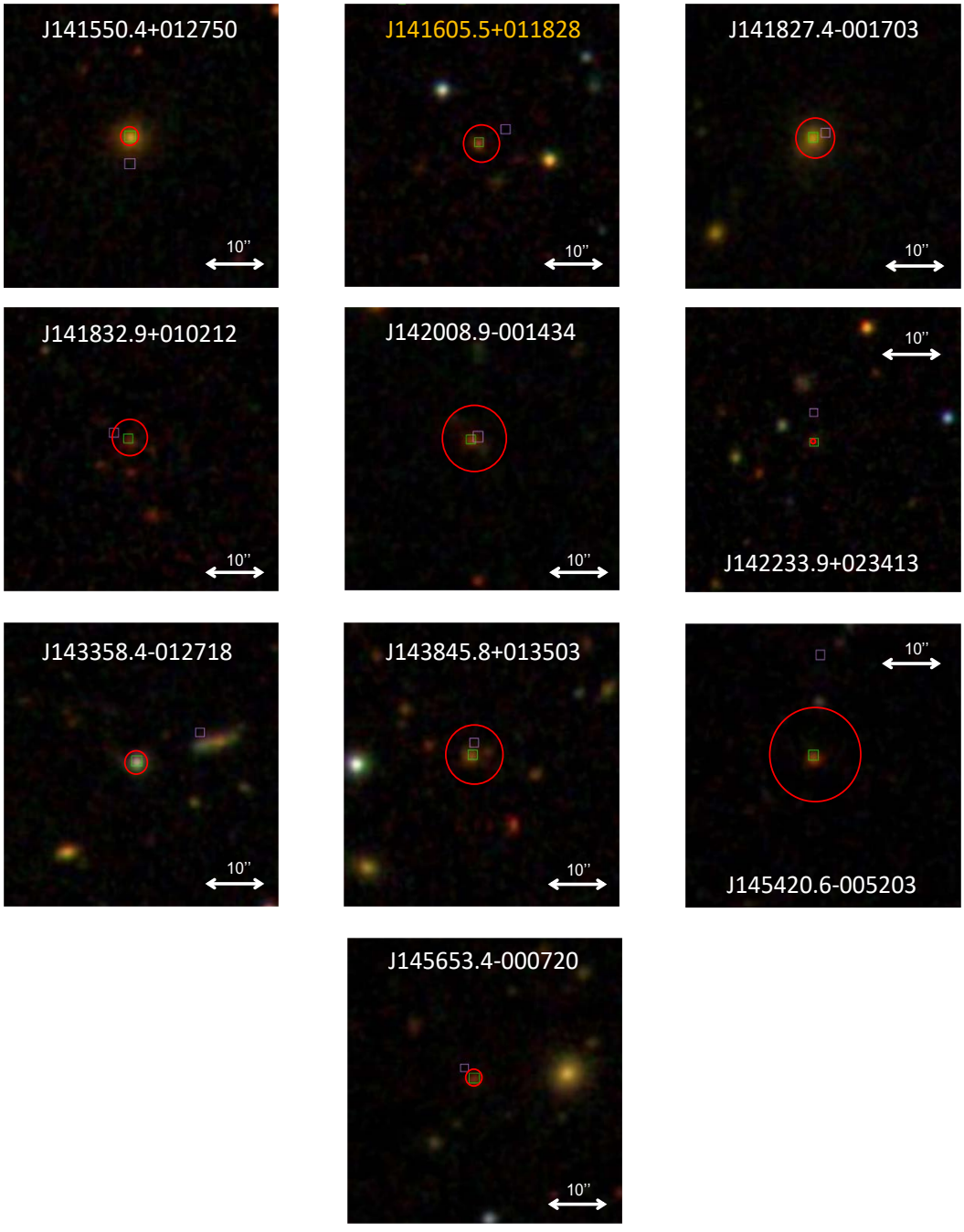


Figure C (cont.)

Appendix D: CNN results for DES strong lenses

We present here the probability scores (P_{lens}) achieved for a sample of DES strong lens candidates presented in Jacobs et al. (2019) and Huang et al. (2020). The model applied was trained with the *clean KiDS-like training data set* in g, r, and i bands.

ID	P_{lens}	Ref	ID	P_{lens}	Ref
DES J000223.6-350716.4	1	[1]	DES J021225.2-085210.8	1	[1]
DES J003119.7-642037.7	1	[1]	DES J022042.3-453003.1	1	[1]
DES J005012.3-465143.6	1	[1]	DES J022416.2-162236.1	1	[2] (Grade B)
DES J010127.8-334319.2	1	[1]	DES J022434.5-000228.0	1	[2] (Grade A)
DES J010158.0-491738.2	1	[1]	DES J023249.9-032325.9	1	[1]
DES J010910.7-045510.1	1	[1]	DES J023307.1-043838.1	1	[1]
DES J012201.0-583718.6	1	[1]	DES J023527.2-481821.0	1	[1]
DES J012431.2-520702.8	1	[1]	DES J023745.4-180102.1	1	[1]
DES J012753.2-453233.9	1	[1]	DES J023929.7-321129.6	1	[1]
DES J012921.3+003916.9	1	[2] (Grade B)	DES J024809.5-395548.3	1	[1]
DES J013322.1-125201.2	1	[1]	DES J024857.6-605403.5	1	[1]
DES J013403.8-200709.0	1	[1]	DES J024958.7-505638.4	1	[1]
DES J013442.4+043350.0	1	[2] (Grade B)	DES J025623.4-270718.6	1	[1]
DES J013650.9-220027.3	1	[1]	DES J025948.2-080446.0	1	[1]
DES J014546.8-354127.3	1	[1]	DES J030945.1-143716.1	1	[1]
DES J014918.8-134904.6	1	[1]	DES J031050.1-174629.3	1	[2] (Grade B)
DES J015319.3-135131.0	1	[1]	DES J031105.6-260104.0	1	[1]
DES J015843.7-130132.9	1	[1]	DES J031710.2-262515.9	1	[1]
DES J020108.0-155117.0	1	[1]	DES J031937.7-575107.9	1	[1]
DES J020304.0-233802.6	1	[1]	DES J032037.0-162422.2	1	[1]
DES J020517.0-012320.5	1	[1]	DES J032251.9-394609.7	1	[1]
DES J020613.5-011417.4	1	[1]	DES J032711.3-324634.2	1	[1]
DES J020706.7-272644.8	1	[1]	DES J032727.2-132622.6	1	[1]

Table D: P_{lens} scores achieved for a sample of DES strong lens candidates presented in [1] Jacobs et al. (2019) and [2] Huang et al. (2020). The grades stand for the level of confidence on the candidates (A>B).

ID	P_{lens}	Ref
DES J032813.6-214019.5	1	[1]
DES J033202.6-132510.2	1	[1]
DES J033845.7-161923.0	1	[1]
DES J034021.4-253330.1	1	[1]
DES J034435.1-444744.0	1	[1]
DES J034713.3-453506.0	1	[1]
DES J034744.5-245431.5	1	[1]
DES J035242.4-382544.9	1	[1]
DES J035346.2-170639.2	1	[1]
DES J040312.8-151257.7	1	[1]
DES J041238.9-195410.9	1	[1]
DES J041538.6-401232.1	1	[1]
DES J041544.5-414346.2	1	[1]
DES J042042.8-542237.4	1	[1]
DES J042438.7-331741.7	1	[1]
DES J043141.6-431740.7	1	[1]
DES J045751.1-433134.2	1	[1]
DES J045847.9-404503.8	1	[1]
DES J050106.7-505447.3	1	[1]
DES J051641.6-441644.5	1	[1]
DES J053142.5-315851.3	1	[1]
DES J053349.3-253654.4	1	[1]
DES J053605.2-533847.1	1	[1]
DES J054358.0-303449.5	1	[1]
DES J055734.4-415950.3	1	[1]
DES J201419.4-575701.5	1	[1]
DES J220044.5-412820.2	1	[1]
DES J221156.8-543839.9	1	[1]
DES J221912.4-434835.1	1	[1]
DES J231200.4-475411.6	1	[1]
DES J232128.4-463049.4	1	[1]
DES J233459.2-640407.0	1	[1]
DES J235519.2-613637.0	1	[1]
DES J032222.6-422421.8	1	[1]
DES J001916.3-413650.6	1	[1]
DES J002228.5+014813.3	1	[2] (Grade B)
DES J010104.0-173203.8	1	[2] (Grade B)
DES J011357.1-292439.2	1	[1]
DES J015909.6-185660.0	1	[1]
DES J022930.5-033836.2	1	[1]
DES J025052.3-552411.7	1	[1]
DES J042930.8-224359.6	1	[1]
DES J053804.6-473513.7	1	[1]
DES J205308.7-504652.4	1	[1]
DES J213156.3-465544.2	1	[1]
DES J014713.9-120648.0	1	[1]
DES J015042.4-024206.5	1	[1]
DES J043303.5-271423.4	1	[1]
DES J025021.2-254513.8	1	[1]
DES J044311.8-202647.7	1	[1]
DES J230521.7-000211.6	1	[1]
DES J014235.0-164817.5	1	[1]
DES J025335.3-354731.5	1	[1]
DES J033628.6-381208.8	1	[1]
DES J224503.8-501725.3	1	[1]
DES J044805.3-580721.3	1	[1]
DES J041013.3-303725.9	1	[1]
DES J002847.0-510849.0	1	[1]
DES J014944.9-380653.7	1	[1]
DES J014504.3-045551.0	1	[1]
DES J031941.3-173404.4	1	[1]
DES J035720.6-595154.4	1	[1]
DES J212447.2-412815.7	1	[1]
DES J031309.8-200631.5	1	[1]
DES J035447.5-242015.0	1	[1]
DES J032242.4-233954.9	1	[1]
DES J043748.1-513628.0	1	[1]
DES J233130.5+003733.4	1	[1]
DES J020929.3-064311.8	1	[1]
DES J013542.8-203335.5	1	[1]
DES J041242.9-264632.3	1	[1]
DES J014106.1-171323.7	1	[1]
DES J013515.5-172415.4	1	[1]
DES J013642.3-194605.3	1	[1]
DES J034745.4-364737.1	1	[1]
DES J015009.1-030438.3	1	[1]
DES J045733.8-354953.3	1	[1]
DES J022014.3-533510.9	1	[1]
DES J225954.8-450448.5	1	[1]
DES J044909.5-291816.4	1	[1]
DES J012533.6-414218.0	1	[1]
DES J222609.3+004142.1	1	[1]
DES J013522.8-423223.5	1	[1]
DES J225403.0-405549.1	1	[1]
DES J020144.7-273942.4	1	[1]
DES J015824.8-003959.4	1	[1]
DES J015930.8-431756.8	1	[1]
DES J051200.8-504122.8	1	[1]
DES J040314.1-150244.9	1	[2] (Grade B)
DES J042934.7-601126.3	1	[1]
DES J014953.7-130609.7	1	[1]
DES J213217.6-430535.1	1	[1]

Table D (cont.)

ID	P_{lens}	Ref
DES J011756.9-242824.6	1	[1]
DES J033459.3-263431.7	1	[1]
DES J021958.0-145645.6	1	[2] (Grade B)
DES J032125.6-153003.5	1	[1]
DES J015153.4-144824.8	1	[2] (Grade A)
DES J042816.3-321800.0	1	[1]
DES J030151.3-442626.0	1	[1]
DES J013718.8-083055.9	1	[1]
DES J014741.0-472640.3	1	[1]
DES J024129.6-594932.9	1	[1]
DES J205615.8-423857.2	1	[1]
DES J011218.1-043458.7	1	[1]
DES J202518.3-420456.9	1	[1]
DES J233734.7-473024.1	1	[1]
DES J010238.3+015856.7	1	[1]
DES J055949.5-354012.5	1	[1]
DES J013354.7-643413.0	1	[1]
DES J024217.3-162441.8	1	[2] (Grade B)
DES J040322.7-543237.1	1	[1]
DES J233607.1-535235.8	1	[1]
DES J014556.3+040229.0	1	[1]
DES J012453.1-144302.6	1	[1]
DES J010210.3-001906.2	1	[2] (Grade B)
DES J030352.3-502351.9	1	[1]
DES J053724.4-464702.4	1	[1]
DES J024216.2-020749.3	1	[1]
DES J004741.6-060611.5	1	[2] (Grade B)
DES J004405.3-044820.9	1	[2] (Grade B)
DES J014134.0-404033.3	1	[1]
DES J001310.1-033545.5	1	[1]
DES J013050.5-160008.5	1	[1]
DES J011301.2-034315.6	1	[2] (Grade B)
DES J033030.8-212711.3	1	[1]
DES J040624.5-264625.0	1	[1]
DES J010510.1-182322.2	1	[2] (Grade B)
DES J221913.2-450450.6	1	[1]
DES J230527.8-444146.6	1	[1]
DES J030920.6-380545.7	1	[1]
DES J230003.2-445423.3	1	[1]
DES J011408.4-361313.5	1	[1]
DES J032801.2-402957.2	1	[1]
DES J023906.5-204718.0	1	[1]
DES J005650.7-411736.8	1	[1]
DES J001542.8-463611.0	1	[1]
DES J043022.2-205110.0	1	[1]
DES J011336.9-185939.1	1	[2] (Grade B)
DES J013415.8-291039.9	1	[1]
DES J001309.6+004003.6	1	[1]
DES J012932.0-114214.0	1	[1]
DES J010530.3-295249.1	1	[1]
DES J015947.0-481812.7	1	[1]
DES J045229.7-354027.2	1	[1]
DES J012611.9-131642.1	1	[1]
DES J050304.8-355306.7	1	[1]
DES J041906.5-295937.8	1	[1]
DES J045332.3-553527.8	1	[1]
DES J022307.7-052758.0	1	[2] (Grade B)
DES J013639.1+000818.1	1	[1]
DES J230136.4-650157.7	1	[1]
DES J003406.4-240525.9	1	[1]
DES J020622.0-205420.2	1	[1]
DES J214052.0-014937.8	1	[1]
DES J010202.9-150028.3	1	[1]
DES J022629.9+020646.9	1	[1]
DES J034805.8-265513.7	1	[1]
DES J023505.9-451054.2	1	[1]
DES J234930.1-511339.0	1	[1]
DES J213758.0-012923.9	1	[1]
DES J034130.8-513044.7	1	[1]
DES J031337.4-361039.9	1	[1]
DES J050921.4-534212.6	1	[1]
DES J005735.2-590518.5	1	[1]
DES J021408.0-020628.4	1	[1]
DES J223421.4-411040.0	1	[1]
DES J215252.9-525006.0	1	[1]
DES J014908.0-313738.3	1	[1]
DES J032028.0-415919.7	1	[1]
DES J020504.9-403828.2	1	[1]
DES J014313.2-501012.8	1	[1]
DES J042909.0-282616.8	1	[1]
DES J001552.1-023045.1	1	[1]
DES J011540.8-352019.4	1	[1]
DES J044129.2-431414.0	1	[1]
DES J035736.8-581053.4	1	[1]
DES J022547.7-150531.9	1	[1]
DES J002457.0-340053.2	1	[1]
DES J031005.4-090314.5	1	[1]
DES J024328.0-214201.9	1	[1]
DES J024314.7-000859.6	1	[2] (Grade B)
DES J045741.3-411524.7	1	[1]
DES J005834.7-520159.6	1	[1]
DES J020651.2-180721.2	1	[1]

Table D (cont.)

ID	P_{lens}	Ref
DES J045008.8-571520.0	1	[1]
DES J015050.0-221615.3	0.999	[1]
DES J223851.0-431730.2	0.999	[1]
DES J024244.3-481136.8	0.999	[1]
DES J020526.1-353947.4	0.999	[1]
DES J021514.3-290925.6	0.999	[1]
DES J010540.5-393926.2	0.999	[1]
DES J021631.0-292006.7	0.999	[1]
DES J033717.2-315213.5	0.999	[1]
DES J033641.8-202111.2	0.999	[1]
DES J014326.9-085021.3	0.999	[1]
DES J035652.5-145758.8	0.999	[1]
DES J014433.3-114211.8	0.999	[1]
DES J033420.2-494040.4	0.999	[1]
DES J030032.7-514444.2	0.999	[1]
DES J013002.9-374457.8	0.999	[1]
DES J010548.0-372542.5	0.999	[1]
DES J030437.7-584647.8	0.999	[1]
DES J034527.9-245918.5	0.999	[1]
DES J024106.6-411507.0	0.999	[1]
DES J021737.2-051329.4	0.999	[1]
DES J044640.0-271447.5	0.999	[1]
DES J023727.3-091308.5	0.999	[1]
DES J041108.0-225635.2	0.999	[1]
DES J010450.9+002514.5	0.999	[2] (Grade B)
DES J210102.0-562947.6	0.998	[1]
DES J211921.5-000946.6	0.998	[1]
DES J004602.0-015628.2	0.998	[1]
DES J031543.3-622035.6	0.998	[1]
DES J051012.8-323205.7	0.998	[1]
DES J012428.5-291856.2	0.998	[1]
DES J042218.2-213245.9	0.998	[1]
DES J043348.7-445605.7	0.997	[1]
DES J050547.0-614950.6	0.997	[1]
DES J023648.1+032130.0	0.997	[1]
DES J233551.9-515217.8	0.996	[1]
DES J011842.6-615613.2	0.996	[1]
DES J051603.2-220847.1	0.996	[1]
DES J004127.1-010747.3	0.996	[2] (Grade B)
DES J030021.6-500128.9	0.996	[1]
DES J021036.8-535152.6	0.995	[1]
DES J210607.2-441153.6	0.995	[1]
DES J001153.1-461421.9	0.995	[1]
DES J013026.5-152012.9	0.994	[1]
DES J030919.7-133212.3	0.994	[1]
DES J012302.4-131321.7	0.994	[2] (Grade B)
DES J041341.1-234422.1	0.994	[1]
DES J025219.9-473237.8	0.994	[1]
DES J012227.3-365426.6	0.994	[1]
DES J040011.3-155247.6	0.994	[2] (Grade B)
DES J224836.4-012334.0	0.993	[1]
DES J044615.0-200142.7	0.993	[1]
DES J045639.7-294621.8	0.993	[1]
DES J011758.7-052717.7	0.993	[1]
DES J052833.4-381141.8	0.992	[1]
DES J033458.4-183838.5	0.992	[1]
DES J004039.5-373249.3	0.992	[1]
DES J030516.7-163613.1	0.991	[1]
DES J010807.1-033446.6	0.991	[2] (Grade B)
DES J012025.8-182001.7	0.989	[1]
DES J041149.7-314454.7	0.989	[1]
DES J030750.5-504204.4	0.988	[1]
DES J032860.0-383309.8	0.988	[1]
DES J025914.1-152148.7	0.987	[1]
DES J212651.1-005827.3	0.986	[1]
DES J202855.8-523118.4	0.986	[1]
DES J025330.8-123903.2	0.986	[1]
DES J005804.9-231713.7	0.986	[1]
DES J015815.3-291218.1	0.985	[1]
DES J033338.5-183708.5	0.985	[1]
DES J014831.5-225127.9	0.983	[1]
DES J031418.6-284156.4	0.982	[1]
DES J051013.2-563754.2	0.980	[1]
DES J051116.8-313432.2	0.980	[1]
DES J032016.1-311917.4	0.980	[1]
DES J004257.3-371858.4	0.980	[1]
DES J012420.6-240147.7	0.978	[1]
DES J005746.8-164444.5	0.978	[2] (Grade B)
DES J010111.8+001002.3	0.977	[2] (Grade B)
DES J234501.2-412610.9	0.977	[1]
DES J041639.6-590852.9	0.976	[1]
DES J012153.1-175538.6	0.976	[2] (Grade B)
DES J010519.6+014456.5	0.975	[1]
DES J015642.7-101100.2	0.974	[1]
DES J054605.1-200025.5	0.973	[1]
DES J010535.3+000700.8	0.972	[2] (Grade B)
DES J033458.6-603450.0	0.970	[1]
DES J022931.0-290816.3	0.970	[1]
DES J025135.3-122001.2	0.964	[1]
DES J212212.4-425060.0	0.964	[1]
DES J022809.1-125252.3	0.964	[2] (Grade A)
DES J235902.2-553805.9	0.963	[1]

Table D (cont.)

ID	P_{lens}	Ref
DES J024604.9-060739.0	0.962	[2] (Grade A)
DES J033410.9-481713.3	0.962	[1]
DES J061312.3-555224.5	0.960	[1]
DES J004613.9-574157.5	0.958	[1]
DES J030348.2-462626.4	0.957	[1]
DES J035418.3-160952.2	0.957	[1]
DES J011646.8-243702.1	0.956	[1]
DES J025243.3-214533.7	0.955	[1]
DES J003727.1-413149.9	0.955	[1]
DES J005828.8-002745.7	0.954	[2] (Grade B)
DES J010605.5+003246.7	0.954	[2] (Grade A)
DES J223233.8-595953.2	0.951	[1]
DES J005212.9-465026.9	0.950	[1]
DES J210504.9-493441.8	0.949	[1]
DES J050417.6-354806.1	0.947	[1]
DES J213047.0+015956.7	0.947	[1]
DES J235846.3-563849.9	0.944	[1]
DES J000729.3-443446.2	0.942	[1]
DES J014905.2-165854.8	0.940	[1]
DES J222052.0-594927.0	0.939	[1]
DES J205158.4-645037.1	0.939	[1]
DES J031435.5-212759.9	0.938	[1]
DES J032629.5-564542.1	0.937	[1]
DES J002510.1-313927.2	0.935	[1]
DES J041809.9-545735.0	0.935	[1]
DES J050747.0-534841.3	0.933	[1]
DES J014656.0-092952.0	0.932	[1]
DES J233448.3-571503.6	0.931	[1]
DES J004659.6+012736.1	0.930	[1]
DES J012503.5-364559.0	0.930	[1]
DES J042225.3-403155.8	0.928	[1]
DES J024228.9-294305.3	0.927	[1]
DES J013250.0-470737.2	0.921	[1]
DES J232346.1-003038.0	0.919	[1]
DES J030918.0-623942.4	0.916	[1]
DES J022140.1-021020.3	0.916	[1]
DES J012052.2-152400.9	0.916	[1]
DES J015452.6-482855.2	0.913	[1]
DES J010146.0+031423.6	0.908	[2] (Grade B)
DES J022501.0-053620.9	0.907	[2] (Grade B)
DES J045901.5-204506.8	0.904	[1]
DES J024449.4-301916.8	0.902	[1]
DES J021050.2-011905.8	0.901	[1]
DES J024551.4-512951.0	0.900	[1]
DES J024303.0-000600.2	0.897	[1]
DES J214006.3-420746.4	0.896	[1]
DES J062500.1-452608.7	0.894	[1]
DES J001718.1+015818.6	0.891	[1]
DES J010159.6-212655.4	0.889	[1]
DES J010659.1-443201.4	0.888	[1]
DES J010315.9+000456.3	0.888	[2] (Grade A)
DES J031127.2-423219.1	0.885	[1]
DES J010726.3+031249.0	0.884	[2] (Grade B)
DES J031741.7-211818.0	0.882	[1]
DES J035649.1-240841.1	0.881	[1]
DES J215410.9+003758.4	0.880	[2] (Grade B)
DES J043123.0-542244.5	0.879	[1]
DES J040205.7-220556.3	0.874	[1]
DES J015148.4-323715.8	0.874	[1]
DES J051407.3-545654.1	0.872	[1]
DES J032903.9-565658.0	0.872	[1]
DES J040025.0-162352.4	0.871	[2] (Grade A)
DES J040058.1-135724.1	0.871	[2] (Grade A)
DES J050655.4-422044.7	0.867	[1]
DES J013823.0-284408.0	0.866	[1]
DES J034919.5-485733.3	0.863	[1]
DES J024553.8-004218.7	0.861	[2] (Grade B)
DES J232243.8-640956.3	0.860	[1]
DES J040715.6-571303.2	0.860	[1]
DES J041142.8-510734.9	0.853	[1]
DES J044229.7-292040.8	0.853	[1]
DES J054224.3-594959.0	0.852	[1]
DES J025340.0-205015.0	0.851	[1]
DES J005616.4-012332.6	0.851	[2] (Grade B)
DES J010306.0-432253.2	0.849	[1]
DES J004827.2+031117.1	0.845	[1]
DES J025051.8-161007.8	0.842	[1]
DES J041644.8-552500.3	0.842	[1]
DES J012441.3-015734.0	0.841	[1]
DES J030729.8-624105.4	0.836	[1]
DES J020234.1-215622.8	0.829	[1]
DES J022844.7-554756.0	0.826	[1]
DES J024440.9-000837.2	0.825	[1]
DES J031638.8-223633.3	0.823	[1]
DES J024116.0-063213.9	0.822	[2] (Grade A)
DES J030221.9-213755.5	0.817	[1]
DES J030647.3-230434.9	0.813	[1]
DES J221205.9-412823.3	0.812	[1]
DES J010228.8-002514.5	0.808	[2] (Grade B)
DES J214915.3-001251.5	0.801	[1]
DES J044248.9-625705.1	0.791	[1]
DES J013327.1-313704.4	0.780	[1]

Table D (cont.)

ID	P_{lens}	Ref
DES J231111.7-454658.1	0.777	[1]
DES J010257.5-291121.8	0.775	[1]
DES J053745.8-471120.3	0.762	[1]
DES J050842.6-274637.4	0.762	[1]
DES J023435.2-622254.5	0.760	[1]
DES J230304.7-511501.8	0.751	[1]
DES J033056.9-522813.0	0.746	[1]
DES J034834.6-130646.9	0.740	[1]
DES J220335.8-640820.7	0.739	[1]
DES J022805.7-051733.7	0.737	[2] (Grade B)
DES J011838.8-052638.8	0.734	[2] (Grade A)
DES J003507.6-252657.6	0.733	[1]
DES J015609.6-102933.4	0.733	[2] (Grade B)
DES J035606.2-560729.4	0.732	[1]
DES J010454.6+000400.5	0.725	[2] (Grade B)
DES J204511.4-634452.5	0.719	[1]
DES J024249.3-003519.7	0.714	[2] (Grade B)
DES J041111.0-481939.6	0.714	[1]
DES J235138.1-545254.0	0.705	[1]
DES J023558.8-060410.6	0.700	[2] (Grade B)
DES J050504.7-461157.5	0.699	[1]
DES J011241.9-550952.9	0.694	[1]
DES J010434.5-003106.6	0.694	[2] (Grade B)
DES J005339.2-050200.9	0.693	[1]
DES J053003.3-610942.5	0.689	[1]
DES J024714.1-375006.1	0.677	[1]
DES J035808.4-295058.7	0.676	[1]
DES J010944.4-143936.0	0.674	[2] (Grade B)
DES J005602.5+030054.7	0.671	[2] (Grade B)
DES J014926.2-382531.4	0.670	[1]
DES J005617.5-022534.3	0.667	[2] (Grade B)
DES J010355.9+001907.3	0.658	[2] (Grade B)
DES J232557.4-005226.7	0.656	[1]
DES J005007.4-183837.7	0.650	[2] (Grade B)
DES J010124.8-170109.1	0.650	[2] (Grade B)
DES J010015.4-001618.5	0.649	[2] (Grade B)
DES J023944.1-462040.8	0.645	[1]
DES J203911.5-545945.4	0.640	[1]
DES J022019.6-041149.6	0.639	[2] (Grade B)
DES J025603.6-121517.9	0.637	[1]
DES J020441.1-150135.8	0.630	[2] (Grade A)
DES J020556.9-020834.7	0.628	[1]
DES J003846.2-293606.8	0.625	[1]
DES J023016.8-312200.8	0.622	[1]
DES J030848.0-210613.9	0.615	[1]
DES J001802.3-454934.4	0.609	[1]
DES J034748.7-315850.3	0.609	[1]
DES J232956.1-532837.7	0.602	[1]
DES J043715.5-642830.1	0.597	[1]
DES J211356.3-011426.5	0.594	[1]
DES J211627.3-594701.8	0.576	[1]
DES J040821.9-535400.9	0.566	[1]
DES J031032.5-464702.3	0.559	[1]
DES J211005.4-563930.6	0.556	[1]
DES J040445.8-143433.6	0.554	[2] (Grade B)
DES J221552.1-013814.0	0.551	[1]
DES J001424.3+004145.5	0.547	[1]
DES J005055.2-172032.5	0.546	[1]
DES J010709.0+011729.0	0.537	[2] (Grade B)
DES J022546.1-073738.4	0.536	[1]
DES J062623.1-573030.1	0.535	[1]
DES J004109.0-004348.9	0.519	[1]
DES J044343.9-622805.7	0.516	[1]
DES J023602.3-512143.8	0.508	[1]
DES J211243.1+000920.8	0.507	[1]
DES J004701.6-290517.6	0.503	[1]
DES J015602.7-641730.0	0.500	[1]
DES J031958.1-475915.5	0.496	[1]
DES J062809.4-502359.8	0.495	[1]
DES J010752.7-142652.8	0.492	[2] (Grade B)
DES J215907.2-430746.1	0.487	[1]
DES J035048.6-482422.3	0.487	[1]
DES J023211.2+001339.2	0.485	[1]
DES J010952.7-333533.2	0.484	[1]
DES J024745.4-003128.9	0.477	[2] (Grade B)
DES J043806.2-322852.2	0.467	[1]
DES J002700.1-041323.6	0.466	[1]
DES J043756.8-650216.3	0.457	[1]
DES J222835.2-465023.8	0.457	[1]
DES J015904.0-341304.3	0.450	[1]
DES J040427.2-171229.5	0.448	[2] (Grade B)
DES J210635.5-540227.9	0.447	[1]
DES J212707.0-514951.0	0.446	[1]
DES J012109.4-243057.3	0.445	[1]
DES J003640.8+004607.3	0.427	[2] (Grade B)
DES J060246.9-452443.2	0.424	[1]
DES J052528.0-442413.5	0.422	[1]
DES J041825.8-612526.9	0.421	[1]
DES J040257.6-525842.4	0.414	[1]
DES J221638.2-441919.8	0.413	[1]
DES J235933.5+020823.6	0.412	[2] (Grade A)
DES J211825.7-431739.6	0.410	[1]

Table D (cont.)

ID	P_{lens}	Ref
DES J032602.9-481235.8	0.408	[1]
DES J031902.2-531819.7	0.406	[1]
DES J055424.3-223806.2	0.404	[1]
DES J010017.3+000009.7	0.401	[2] (Grade B)
DES J213906.0-425132.5	0.400	[1]
DES J011607.3-281224.9	0.399	[1]
DES J010236.7-000546.7	0.397	[2] (Grade B)
DES J031103.5-131402.8	0.397	[2] (Grade B)
DES J000316.4-334804.3	0.393	[1]
DES J211209.0-014524.5	0.393	[2] (Grade A)
DES J033906.5-583016.9	0.392	[1]
DES J045443.7-623400.4	0.390	[1]
DES J031830.0-194227.6	0.389	[1]
DES J010651.6-625813.8	0.388	[1]
DES J025144.2-061329.4	0.388	[1]
DES J054627.5-332922.7	0.381	[1]
DES J220106.5-604702.7	0.380	[1]
DES J051300.2-384743.2	0.376	[1]
DES J034802.3-214503.0	0.375	[1]
DES J034948.0-150008.4	0.369	[1]
DES J000644.6-442950.5	0.366	[1]
DES J020206.6-244503.8	0.364	[1]
DES J224844.0-443150.9	0.364	[1]
DES J062019.0-562858.3	0.361	[1]
DES J045536.7-253046.1	0.356	[1]
DES J023953.1-013456.0	0.355	[1]
DES J062415.8-470942.0	0.353	[1]
DES J032216.4-523440.4	0.349	[1]
DES J061227.5-392036.3	0.349	[1]
DES J030505.5-102411.7	0.340	[1]
DES J050849.3-214430.7	0.333	[1]
DES J221857.6-450425.6	0.325	[1]
DES J055311.2-285336.9	0.324	[1]
DES J214454.4-414952.6	0.316	[1]
DES J033356.0-144008.9	0.316	[1]
DES J025925.9-520642.9	0.315	[1]
DES J024706.5-591731.7	0.305	[1]
DES J052221.1-603624.0	0.305	[1]
DES J015752.3-531155.2	0.290	[1]
DES J053444.8-534716.3	0.281	[1]
DES J022709.0-471855.8	0.256	[1]
DES J225506.3-412320.2	0.256	[1]
DES J222629.0-463605.4	0.248	[1]
DES J232510.9-411124.9	0.236	[1]
DES J230824.7-021213.6	0.233	[1]
DES J030745.3-181108.1	0.230	[1]
DES J001332.8-423929.1	0.227	[1]
DES J014252.9-183115.8	0.214	[1]
DES J011241.1-190244.3	0.213	[1]
DES J015216.3-583842.3	0.213	[1]
DES J002606.7-550432.9	0.188	[1]
DES J011005.8-023230.1	0.182	[2] (Grade B)
DES J230822.2-021131.7	0.179	[1]
DES J214546.4-430649.3	0.176	[1]
DES J052423.7-272111.4	0.171	[1]
DES J030416.1-492126.1	0.162	[1]
DES J032407.0-355611.0	0.138	[1]
DES J024700.1-443211.0	0.138	[1]
DES J024803.4-033144.9	0.126	[1]
DES J022057.7-383303.3	0.110	[1]
DES J001030.4-431514.9	0.108	[1]
DES J211723.6-005651.9	0.085	[1]
DES J024524.8-530145.4	0.058	[1]
DES J011225.8+022238.3	0.052	[2] (Grade A)
DES J212512.0-650426.7	0.045	[1]
DES J010131.3+003412.0	0.028	[2] (Grade B)
DES J022534.0-050211.4	0.021	[2] (Grade B)
DES J050600.5-204900.5	0.018	[1]
DES J034204.6-535514.1	0.015	[1]
DES J054302.1-375237.2	0.013	[1]
DES J030553.4-145252.3	0.009	[2] (Grade B)
DES J002134.0-404001.8	0	[1]

Table D (cont.)

Resumen de la tesis en castellano

8.1 Introducción

La segunda mitad del siglo XX fue testigo del florecimiento como ciencia de la Cosmología observacional. La detección de la radiación del fondo cósmico de microondas (RFCM) en 1964, junto con el descubrimiento de su estructura y anisotropías en 1992; las cantidades observadas de elementos ligeros (hidrógeno, deuterio, helio, y litio) en el universo; su expansión acelerada, medida y confirmada en 1998; y su estructura a gran escala constituyen las principales evidencias observacionales que apoyan el modelo ΛCDM. Este modelo paramétrico, a menudo referido como el modelo estándar de la Cosmología, es actualmente nuestra mejor explicación para todas las observaciones del Universo. De acuerdo con este modelo, en el universo existen tres componentes principales: primero, la energía oscura, que parece comportarse como una constante cosmológica (Λ); segundo, la materia oscura; y por último la materia ordinaria. Según las últimas medidas de las anisotropías de la RFCM (Planck Collaboration, 2020), sólo el 4.9% de la densidad de energía en el universo corresponde a matera bariónica. La conocida como materia oscura, materia no bariónica débilmente interactuante, representa el 26.4% mientrás que el 68.7% restante se identifica como energía oscura.

La expansión acelerada es uno de los fenónemos más misteriosos del universo, consecuencia de lo poco que sabemos sobre la energía y la materia oscura. Y por tanto, todavía estamos lejos de tener un conocimiento completo de nuestro universo. Por esta razón, se están realizando numerosos esfuerzos por desarrollar teorías y modelos que concuerden con todas las pruebas que sostienen el modelo ΛCDM, pero que arrojen luz sobre aquellas observaciones que no alcanza a explicar. Revelar la naturaleza de estos dos componentes oscuros del universo es uno de los desafíos de la Cosmología para los próximos años. Esta desafiante tarea requiere poder sondear el universo primitivo con gran detalle. En este contexto, el desarrollo de técnicas de detección de objetos astronómicos débiles con alto corrimiento al rojo (*redshift*) se convierte en una cuestión de gran importancia.

8.2 Detección multifrecuencia de fuentes extragalacticas

Durante las últimas décadas, los avances en la sensibilidad de las observaciones (especialmente en la parte infrarroja del espectro) y los progresos en el procesamiento de datos nos han permitido sondear el universo a alto redshift en mayor profundidad. La observación directa de galaxias en el rango $z \sim 1-10$ nos da la oportunidad de estudiar la historia de la formación de galaxias y estrellas en diferentes épocas cósmicas (ver, por ejemplo, de Zotti et al., 2010; Eales, 2015). Sin embargo, a pesar del constante aumento de la sensibilidad de los detectores y del diámetro de los telescopios, las observaciones del universo distante aún tienen un flujo limitado, siendo visibles sólo aquellos objetos que están por encima del umbral de detección del instrumento. En un universo en el que prevalece la ley del cuadrado inverso, un límite de flujo implica que las galaxias con mayor redshift accesibles a cualquier observatorio estarán entre sus objetos detectados más débiles. Esta situación se ve aliviada para las fuentes seleccionadas en el rango submilimétrico del espectro electromagnético gracias a la fuerte corrección negativa K, que lleva a que las galaxias con alto redshift sean relativamente fáciles de detectar en longitudes de onda submilimétricas en comparación con sus equivalentes a bajo redshift (Blain and Longair, 1993). Además, las posibles alineaciones fortuitas de objetos de fondo con lentes en primer plano pueden ampliar los límites aún más, al mejorar el flujo de objetos que no podrían detectarse de otra manera. Pero incluso con la ayuda de la corrección K y del efecto lente gravitacional, las técnicas de procesamiento de señales son una herramienta fundamental para llegar a las galaxias más débiles y distantes. Esto es particularmente cierto para los rangos de microondas e infrarrojo lejano del espectro electromagnético, donde las fluctuaciones del fondo infrarrojo cósmico crean un ruido de confusión cuyo nivel es comparable con la densidad de flujo de las galaxias típicas a alto refshift.

8.2.1 Técnicas de filtrado y detección

El filtrado es un tipo de procesamiento de señal que permite seleccionar las frecuencias deseadas, o eliminar las que no interesan, de unos datos brutos para favorecer la detección de una determinada señal. En el formalismo matemático, un filtro viene representado por la siguiente operación:

$$L : f(x) \rightarrow g(x) = Lf(x) \tag{8.12}$$

donde f son los datos de entrada, L representa al filtro, g es la señal de salida obtenida

tras el filtrado, y x es la variable independiente. Un filtro lineal es aquel en el que la señal filtrada es una función lineal de los datos de entrada. Y se dice que un filtro es homogéneo, o invariante en el tiempo, si su señal de salida se retrasa un cierto tiempo τ cuando la señal de entrada también se retrasa ese mismo tiempo. La homogeneidad de un filtro puede expresarse como $g(x-\tau) = Lf(x-\tau)$. La mayoría de los filtros utilizados en un amplio abanico de campos científicos son lineales y homogéneos. Cualquier filtro lineal invariante en el tiempo se puede caracterizar por su respuesta de impulso. Esto significa que para cualquier señal de entrada, la salida se puede obtener en términos de la entrada y la respuesta al impulso. La función de respuesta al impulso de un filtro es su salida cuando recibe una breve señal de entrada, llamada impulso. Este impulso generalmente se modela para señales continuas como una delta de Dirac, cuyo valor en x es obtenido por la integral:

$$f(x) = \int f(u)\delta(x-u)du \tag{8.13}$$

La linealidad y continuidad del filtro L implican que

$$Lf(x) = \int f(u)L\delta(x-u)du \tag{8.14}$$

Denotando la respuesta al impulso del filtro L como $h(x) = L\delta(x)$, podemos reescribir Eq. (8.14) de la siguiente forma:

$$Lf(x) = \int_{-\infty}^{\infty} f(u)h(x-u)du = h \otimes f \tag{8.15}$$

donde $\otimes$ representa una convolución. Esta ecuación muestra que un filtro lineal homogéneo equivale a una convolución con la respuesta al impulso h del filtro. El teorema de la convolución establece que la convolución de dos funciones, h y f, en un punto de coordenadas x se puede expresar como:

$$Lf(x) = h(x) \otimes f(x) = \int_{-\infty}^{\infty} \hat{h}(q)\hat{f}(q)e^{-iqx}dq \tag{8.16}$$

donde $\hat{h}(q)$ y $\hat{f}(q)$ son las correspondientes transformadas de Fourier de $h(x)$ y $f(x)$, respectivamente, de acuerdo con la siguiente convención:

$$\hat{f}(q) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{iqx}dx, \quad f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(q)e^{-iqx}dq \tag{8.17}$$

La transformada de Fourier $\hat{h}(q)$ de la respuesta al impulso $h(x)$ también se conoce como función de transferencia del filtro. Considerándolo todo, vemos que el proceso de filtrar una imagen con un filtro lineal homogéneo equivale a multiplicar la transformada de Fourier de la imagen por esta función de transferencia. Esta función de transferencia puede, por tanto, considerarse como un dispositivo selectivo de frecuencia, en el sentido de los modos de Fourier. Esta propiedad hace que trabajar en el espacio de Fourier sea la mejor opción. Además, las señales que queremos detectar pertenecen a fuentes compactas o puntuales, cuyos perfiles son bien conocidos. Los modos de Fourier de estas fuentes compactas se obtienen fácilmente de sus perfiles. Por estas razones, el filtrado ha demostrado ser muy adecuado para la detección de fuentes compactas incrustadas en un fondo ruidoso. Solo tenemos que diseñar funciones de transferencia capaces de reducir la contribución de las frecuencias responsables del ruido, a la vez que conserven las frecuencias características correspondientes a las fuentes compactas.

El filtrado adecuado de imágenes astronómicas debe comenzar por ser continuo en el espacio de Fourier. El siguiente paso es encontrar un filtro que mejore la detección de fuentes compactas incrustadas en un fondo ruidoso, reduciendo dicho ruido y conservando la señal. El criterio más utilizado para lograr esto busca aumentar tanto como sea posible la relación señal/ruido (S/R) de las fuentes que tratamos de detectar. Supongamos una señal s con amplitud A en la posición x_0 incrustada en un fondo ruidoso de dispersión σ. El cociente S/R de esta señal se define como la relación entre su amplitud y la desviación estándar de todo el campo:

$$S/R = \frac{s(x_0)}{\sigma} = \frac{A}{\sigma} \tag{8.18}$$

Obviamente, el cociente S/R en la imágen filtrada cambia dependiendo del filtro utilizado.

Los métodos monofrecuencia estándar para la detección de fuentes puntuales en la RFCM y el infrarrojo lejano se basan en técnicas de ondículas (*wavelets,* Vielva et al., 2003; Barnard et al., 2004; González-Nuevo et al., 2006) o en el filtro 'ajustado' (*matched filter,* Tegmark and de Oliveira-Costa, 1998; Herranz et al., 2002; Barreiro et al., 2003; López-Caniego et al., 2006, vease también Herranz and Vielva (2010) para una revisión). Las wavelets son muy adecuadas para la detección de fuentes compactas debido a su capacidad para descomponer el comportamiento de la señal en distintas escalas, mientras que el filtro ajustado es el detector lineal óptimo porque proporciona la máxima amplificación S/R para una fuente con una forma conocida (generalmente la función de dispersión de punto, o PSF en lo sucesivo, del telescopio) incrustada en ruido estadísticamente homogéneo y correlacionado espacialmente. Por defecto, estas técnicas solo se aplican a imágenes

del cielo en una sóla frecuencia. Incluso para observatorios multifrecuencia como *Herschel* Space Observatory (Pilbratt et al., 2010) o *Planck* (Tauber et al., 2010), los procedimientos de detección estándar han producido catálogos de fuentes individuales para cada banda de frecuencia (ver, por ejemplo, Planck Collaboration, 2011, 2014, 2016; Maddox et al., 2018). El siguiente paso lógico, por tanto, es potenciar la señal de fuentes débiles combinando las detecciones de diferentes bandas en una sola detección, es decir, "detección multifrecuencia". La mayoría de los algoritmos ciegos de separación de componentes que se utilizan para componentes difusos en el rango de microondas e infrarrojo lejano no pueden lidiar con la alta diversidad de comportamientos espectrales asociados a las diferentes poblaciones de fuentes compactas extragalácticas (ver, por ejemplo, Leach et al., 2008). Sin embargo, en los últimos años se han propuesto varias técnicas de detección de fuentes compactas multifrecuencia en la literatura (Herranz and Sanz, 2008; Herranz et al., 2009; Lanz et al., 2010, 2013; Planck Collaboration, 2018). En particular, si se conoce el perfil espacial y la distribución espectral de energía (SED) de las fuentes, y si se conoce también el espectro de potencias cruzado, o se puede estimar a partir de los datos, el método de detección lineal óptimo es el multifiltro ajustado (*matched multifilter* o MMF, Herranz et al., 2002). Esta generalización supera al filtro ajustado monofrecuencia en términos de relación S/R.

8.2.2 El multifiltro ajustado

En primer lugar, supongamos que se toman imágenes a N frecuencias diferentes del mismo área del cielo. Para simplificar, también asumimos que en las imágenes solo hay una fuente puntual, de amplitud A_ν en cada frecuencia, centrada en el origen de coordenadas. La señal en estas imágenes se puede describir como:

$$y_\nu(\mathbf{x}) = f_\nu s_\nu(\mathbf{x}) + n_\nu(\mathbf{x}) \tag{8.19}$$

donde y_ν es la señal total en el píxel $\mathbf{x}$, f_ν es la dependencia frecuencial de la fuente puntual, s_ν es la contribución de la fuente puntual a la señal total, y n_ν representa el ruido de fondo en cada píxel, que alberga tanto el ruido instrumental como el perteneciente a otras contribuciones detectadas. Estas variables se miden para cada una de las frecuencias de observación $\nu = 1, ..., N$ consideradas.

Antes de obtener ninguna imagen, la señal de esta fuente puntual llega a un detector. La resolución angular de este detector es mayor que el tamaño angular intrínseco de la fuente puntual. Luego, su señal se convoluciona, en cada frecuencia de observación ν, con el haz de antena correspondiente. Suponemos, de nuevo por simplicidad, que los haces de

antena pueden describirse bien mediante funciones gaussianas bidimensionales simétricas. La señal de la fuente puntual puede entonces escribirse como:

$$s_\nu(x) = A_\nu \tau_\nu(x) \tag{8.20}$$

donde $\tau_\nu(x)$ es el perfil espacial de la fuente, y $x = |\mathbf{x}|$ debido al supuesto de que las antenas consideradas son simétricas. El ruido de fondo $n_\nu(\mathbf{x})$ se puede modelar como un campo aleatorio homogéneo e isotrópico con valor promedio igual a cero. El espectro de potencias cruzado $\mathbf{P} = (P_{\nu_1\nu_2})$ entre los ruidos en dos de estas frecuencias, ν_1 and ν_2, es definido por:

$$\langle n_{\nu_1}(\mathbf{q}) n^*_{\nu_2}(\mathbf{q'})\rangle = P_{\nu_1\nu_2}\delta^2_D(\mathbf{q}-\mathbf{q'}) \tag{8.21}$$

siendo $n_\nu(\mathbf{q})$ la transformada de Fourier de $n_\nu(\mathbf{x})$, δ^2_D la delta de Dirac bidimensional, y el símbolo $*$ indica conjugación compleja.

El conjunto de N filtros lineales invariantes en el tiempo ψ_ν que componen el MMF son aquellos que producen los siguientes mapas:

$$\omega_{\psi_\nu}(\mathbf{b}) = \int y_\nu(\mathbf{x})\psi_\nu(\mathbf{x};\mathbf{b})d\mathbf{x} = \int y_\nu(\mathbf{q})\psi_\nu(\mathbf{q})e^{-i\mathbf{q}\cdot\mathbf{b}}d\mathbf{q} \tag{8.22}$$

donde $\mathbf{b}$ define una traslación desde el origen de coordenadas, $\omega_\nu(\mathbf{b})$ representa cada imagen filtrada en la frecuencia ν en la posición $\mathbf{b}$, y $y_\nu(\mathbf{q})$ y $\psi_\nu(\mathbf{q})$ son las transformadas de Fourier de $y_\nu(\mathbf{x})$ y $\psi_\nu(\mathbf{x})$, respectivamente. El primer paso en este método multifrecuencia es filtrar cada imagen y_ν con uno de estos filtros lineales ψ_ν. En el siguiente paso, todos los mapas filtrados ω_{ψ_ν} visto en la Eq. (8.22) se combinan de acuerdo con:

$$\omega_\psi(\mathbf{b}) = \sum_\nu \omega_{\psi_\nu}(\mathbf{b}) \tag{8.23}$$

Este mapa filtrado total es una imagen en la que la señal de la fuente se ha realzado mientras que el ruido se ha reducido significativamente. Los dos requisitos necesarios para garantizar que el campo filtrado sea óptimo para la detección de fuentes puntuales son:

1. El filtro debe ser un estimador *insesgado* de la amplitud de la fuente. Esto significa que, después de filtrar la imagen, el valor de amplitud A de la fuente debe recuperarse, en promedio, tras muchas realizaciones, en la posición de la fuente. Para un caso concreto, incluso después del filtrado, siempre habrá alguna contribución del ruido.

2. El filtro debe ser un estimador de *máxima eficiencia*. Esto significa que la varianza σ_ψ^2 alrededor del valor promedio recuperado de A tiene que ser lo más pequeña posible.

Como dijimos anteriormente, el filtro que buscamos se llama multifriltro ajustado, y se introdujo por primera vez para la detección de cúmulos Sunyaev-Zel'dovich (SZ) en mapas multifrecuencia (Herranz et al., 2002). El MMF es el método de detección lineal óptimo cuando la dependencia de la frecuencia y el perfil espacial de las fuentes se conocen, o se consideran conocidos, y el espectro de potencias cruzado del ruido se conoce o se puede estimar a partir de los datos. En el espacio de Fourier, el MMF se puede escribir de la siguiente manera:

$$\begin{aligned} \boldsymbol{\Psi}(q) &= \alpha\, \mathbf{P}^{-1}\mathbf{F}, \\ \alpha^{-1} &\equiv \int \mathbf{F}^t\, \mathbf{P}^{-1}\, \mathbf{F}\, dq, \\ \sigma^2 &= \int \boldsymbol{\Psi}^t\, \mathbf{P}\, \boldsymbol{\Psi}\, dq = \alpha \end{aligned} \tag{8.24}$$

donde $\boldsymbol{\Psi}(q) = [\psi_\nu(q)]$ es el vector columna de los filtros; $\mathbf{F} = [f_\nu \tau_\nu]$, siendo f_ν la dependencia frecuencial y τ_ν el perfil de la fuente en cada frecuencia ν; $\mathbf{P}^{-1}$ es la matriz inversa del espectro de potencias cruzado $\mathbf{P}$; y σ^2 es la varianza de la imagen filtrada producida. En la Eq. (8.24) y en la siguiente discusión, $q \equiv |\mathbf{q}|$ es el módulo del vector de onda en el espacio de Fourier, puesto que asumimos perfiles circularmente simétricos para las fuentes. Y dado que el espectro de potencias cruzado solo depende del módulo q, todas las fórmulas se pueden expresar en términos de q en lugar de usar el vector completo. Finalmente, α en la Eq. (8.24) se puede interpretar como la normalización requerida para garantizar que los filtros ψ_ν sean estimadores insesgados de la densidad de flujo de las fuentes en estudio. Más detalles pueden encontrarse en Herranz et al. (2002) y Lanz et al. (2010, 2013).

El MMF se ve reducido a una matriz diagonal de N filtros ajustados, uno en cada frecuencia, si el ruido en los mapas considerados no presenta una correlación entre las diferentes frecuencias. Por lo tanto, un enfoque multifrecuencia propiamente dicho solo puede llevarse a cabo si los mapas considerados muestran correlaciones entre ellos. El comportamiento espectral f_ν de las fuentes no es conocido a priori solo con la información de las imágenes. Por lo tanto, es necesario encontrar alguna forma de modelar esta desconocida dependencia frecuencial. Lanz et al. (2010) mostró que el MMF puede generalizarse para el caso en el que esta dependencia no se conoce. A pesar de esto, en este trabajo hemos incorporado una SED específica al MMF con el fin de derivar estimaciones fotométricas del redshift de las fuentes en estudio. Para abordar este problema, es muy útil

reescribir el vector $\mathbf{F} = [f_\nu \tau_\nu]$ en la forma matricial $\mathbf{F} = \mathbf{T}(q)\boldsymbol{f}(\nu)$, siendo la matriz diagonal $\mathbf{T}(q) = [\tau_1(q),..., \tau_N(q)]$ y el vector de dependencia frecuencial $\boldsymbol{f}(\nu) = [f_\nu]$. Gracias a ello, podemos incluir toda la dependencia de q en la matriz $\mathbf{T}$, separándola completamente de la dependencia en ν. De esta manera, la Eq. (8.24) se puede reescribir como:

$$\boldsymbol{\Psi}_f(q) = \alpha_f \mathbf{P}^{-1}\mathbf{T}\boldsymbol{f}, \quad \alpha_f^{-1} = \int \boldsymbol{f}^t \mathbf{T}\mathbf{P}^{-1}\mathbf{T}\boldsymbol{f} dq = \boldsymbol{f}^t \mathbf{H}\boldsymbol{f} \tag{8.25}$$

donde matriz $\mathbf{H} = \int \mathbf{T}\mathbf{P}^{-1}\mathbf{T}dq$, y nos basamos en el hecho de que $\mathbf{T}^t = \mathbf{T}$ y el vector $\boldsymbol{f}$ no depende de q. Esta reformulación del MMF es muy conveniente para su posterior implementación. La parte del filtrado que conlleva más tiempo es el cálculo de las matrices $\mathbf{P}$ y $\mathbf{T}$, ya que deben calcularse para todos los valores de q. En el caso que estamos considerando en este trabajo, la única cantidad que varía durante el proceso de maximización es el redshift de la fuente que queremos estimar. Esto nos permite calcular las integrales de la matriz $\mathbf{H}$ solo una vez para cada conjunto de imágenes de la fuente considerada. La estimación de la amplitud A de una fuente puntual presente en un conjunto de imágenes después de aplicar estos filtros viene dada por:

$$A_{\Psi_f} = \omega_{\Psi_f}(0) = \alpha_f A \boldsymbol{f}^t \mathbf{H}\boldsymbol{f} \tag{8.26}$$

8.3 Aplicación del multifiltro ajustado a fuentes de H-ATLAS

En esta tesis hemos extendido el trabajo previo realizado con la técnica del multifiltro ajustado a las imágenes astronómicas en el rango submilimétrico proporcionadas por *Herschel*-ATLAS. Esta técnica multifrecuencia permite, por un lado, aumentar la significancia estadística y el cociente S/R de las detecciones y, por otro lado, proporciona una estimación fotométrica de los redshifts de las fuentes. Este enfoque se basa en el modelado semi-analítico del pico de emisión térmica de la SED de galaxias polvorientas dado por Pearson et al. (2013) (Eq. 2.1). El uso de plantillas para modelar la SED no es estrictamente necesario para utilizar el MMF, pero resulta muy útil para la estimación del redshift. Hemos limitado nuestro análisis a las tres bandas cubiertas por el instrumento SPIRE, centradas en 250, 350 y 500 μm.

Hemos probado nuestra técnica tanto con simulaciones realistas (Fig. 2.1) como con fuentes de H-ATLAS para las que se conocen con precisión sus redshifts (Fig. 2.2), concluyendo que el multifiltro ajustado, en efecto, conduce a una S/R mejorada con respecto a la detección en una única frecuencia, y que los redshifts fotométricos estimados son rel-

ativamente precisos. Esto es especialmente cierto para aquella fuentes de H-ATLAS con un comportamiento espectral cercano a la SED utilizada en el rango $0.8 < z < 4.3$. Fuera de este rango, hemos estudiado el sesgo en la estimación de los redshifts fotométricos del MMF, y hemos proporcionado factores de corrección hasta $z \sim 7.0$.

En total, 410,997 fuentes puntuales de H-ATLAS (Valiante et al., 2016; Maddox et al., 2018) se han vuelto a analizar con este filtro multifrecuencia. Se ha llevado a cabo una selección de fuentes de H-ATLAS a alto redshift, para las que el multifiltro ajustado funciona mejor, en base a criterios en el flujo, el color, el cociente S/R, y la concordancia entre la fotometría observada y el modelo de SED usado (ver Fig. 2.4). Como resultado tenemos 607 fuentes divididas en dos muestras:

- Una muestra de 370 objetos brillantes a alto redshift, con $\bar{z} = 2.13$ y $\sigma_z = 0.65$ (ver Tabla B.1) Esta muestra contiene 62 de los 80 candidatos a ser sistermas lensados identificados en Negrello et al. (2017), e incluye 17 de las 20 galaxias con lentes fuertes confirmadas. Se ha logrado una mejora promedio del 76% en la S/R para esta muestra con el MMF en comparación con la obtenida en la banda de 500 μm. Además, se ha obtenido una mejora promedio del 16% y una ligera mejora del 0.2% para las bandas de 350 μm y 250 μm, respectivamente (ver Fig. 2.9). Mediante la correlación cruzada de esta muestra a alto z con un catálogo de galaxias a bajo redshift de SDSS DR14, encontramos 40 candidatos potenciales para ser sistemas con lentes gravitacionales. En la Fig. C se muestran imágenes de estos candidatos, algunas de las cuáles revelan sobredensidades de galaxias a $z < 1$ que podrían indicar la presencia de grupos de galaxias actuando como lentes.
- Una segunda muestra de 237 fuentes a alto redshift, con $\bar{z} = 4.62$ y $\sigma_z = 0.71$, candidatos a ser "500 μ-risers" (débiles a 250 y 350 μm, pero brillantes a 500 μm) que antes estaban cerca del límite de detección de H-ATLAS pero ahora están confirmados con el MMF como detecciones con alta significancia (ver Tabla B.2). Hemos logrado mejoras promedio de 25%, 55% y 76% en las relaciones S/R para las bandas de 500, 350 y 250 μm, respectivamente (ver Fig. 2.11). Esto refleja claramente que es en este tipo de objetos débiles donde nuestro método MMF alcanza su máximo potencial en términos de mejora de S/R.

El multifiltro ajustado ha demostrado ser capaz de incrementar la sensibilidad de los objetos más débiles y mejorar su fotometría. También es capaz de devolver redshifts fotométricos más robustos que los que se obtendrían usando los flujos del catálogo de H-ATLAS, como lo demuestran los valores de la media μ y mediana $\mu_{1/2}$ mostrados en las Tablas 2.1 y 2.3.

La ejecución de una búsqueda no ciega, utilizando las posiciones ya conocidas de fuentes en el catálogo de H-ATLAS, nos permite comparar directamente nuestro método con la rutina de detección de H-ATLAS. Sin embargo, dado que, al final, la imagen de detección de H-ATLAS simplemente se reduce al canal de 250 μm (Valiante et al., 2016), en el futuro planeamos aplicar este método realizando una búsqueda ciega sobre todo el campo de visión de H-ATLAS. De ese modo podremos explotar la información de las otras longitudes de onda y ser capaces de detectar fuentes débiles que ahora estaríamos perdiendo.

El trabajo aquí realizado también podría ampliarse en el futuro centrándose únicamente en las fuentes ya confirmadas. Uno de los posibles análisis o mejoras más profundas sería repetir la ejecución del MMF utilizando otras SED diferentes del modelo definido en Pearson et al. (2013). Por ejemplo, SEDs derivadas de las galaxias polvorientas ultraluminosas Arp200; SMM J2135-0102; "The Cosmic Eyelash" a $z = 2.3$ (Ivison et al., 2010) o H-ATLAS J142413.9 + 022304 alias G15.141 en $z = 4.23$ (Cox et al., 2011), las cuales son representaciones características de las galaxias submilimétricas locales. Este procedimiento nos mostraría cuánto cambian los redshifts y las densidades de flujo según la SED utilizada, y nos ayudaría a seleccionar fuentes de una manera más robusta. Este análisis adicional también permitiría comprobar si el sesgo metodológico observado en las simulaciones se debe a la forma particular de la SED empleada. No obstante, pensamos que el modelo utilizado aquí es la mejor opción, ya que no es una SED derivada de una sola fuente, sino un modelo empírico basado en 40 objetos de H-ATLAS, que ya ha sido utilizado en varios trabajos repvios (Eales, 2015; Ivison et al., 2016; Bianchini et al., 2016, 2018; Negrello et al., 2017; Fudamoto et al., 2017; Bakx et al., 2018; Donevski et al., 2018).

8.4 Distribución de materia oscura en cúmulos de galaxias

Como hemos dicho, la materia oscura constituye uno de los principales componentes del universo. A pesar de ello, su naturaleza aún nos es desconocida. Lo que sí sabemos es que no puede estar formada por ninguna de las patículas ordinarias ya descubiertas, es fría, no relativista y debilmente interactuante. Fueron las observaciones de las curvas de rotación de diversas galaxias las que llevaron a la conclusión de que un importante porcentaje de la masa de las galaxias se debe un componente no luminoso, bautizado como materia oscura, puesto que dichas curvas no podían explicarse teóricamente sólo teniendo en cuenta la materia ordinaria asociada a las estrellas.

La masa de las galaxias está dominada por materia oscura, y la evidencia más reveladora que tenemos de su existencia se debe a su interacción gravitatoria. Por lo tanto, una de

las formas de estudiar las propiedades de la materia oscura es a través de esta interacción, siempre que la concentración de materia oscura sea lo suficientemente grande. Los cúmulos de galaxias son una excelente opción para ello (Kravtsov and Borgani, 2012). Los cúmulos son las estructuras más masivas en equilibrio dinámico y, por lo tanto, son sondas altamente sensibles y útiles de la formación de estructuras gravitacionales y de la evolución de las galaxias a lo largo de la historia del universo. Estimar el perfil de masa en la zona más central de los cúmulos es de particular interés, ya que esta altamente densa región contiene las mayores densidades de materia oscura. Estas medidas pueden ser útiles para acotar las propiedades físicas de la materia oscura (Arabadjis et al., 2002; Markevitch et al., 2004; Serra and Domínguez Romero, 2011), y para distinguir entre materia oscura y teorías alternativas de la gravedad (Clowe et al., 2006). Si la materia oscura tiene una sección eficaz de interacción pequeña, o si la materia oscura es ultraligera, es en estas regiones donde se esperaría ver desviaciones de los modelos que predicen una materia oscura fría no colisionante (Rocha et al., 2012; Kaplinghat et al., 2013, 2014). Además, si la materia oscura se aniquila o interactúa consigo misma, se espera que la densidad en la región central de los cúmulos de galaxias sea sensible a estas interacciones. En particular, si la sección eficaz de auto-interacción es significativamente mayor que 1 $\mathrm{cm^2 g^{-1}}$, el perfil de masa debería aplanarse en la región más central y, si el cúmulo está fusionandose, deberían poder observarse dos picos separados, correspondientes a la contribución de la galaxia más brillante del cúmulo (BCG) y al pico de la distribución de materia oscura.

Los perfiles de masa central en los cúmulos de galaxias pueden inferirse de varias formas, cada una de las cuales explora un rango de radios diferente. Estos métodos incluyen el estudio de la cinemática estelar, del efecto lente gravitatoria, del efecto SZ, y de la emisión en rayos X, los cuales cubren un rango en distancia de 10 kpc a 1 Mpc (Newman et al., 2009; Umetsu et al., 2011; Hogan et al., 2017; Andrade et al., 2019). El estudio de la desviación de la luz en presencia de objetos masivos, y los fenómenos resultantes de ello, es lo que se conoce como efecto lente gravitatoria (*gravitational lensing*). Las estimaciones más precisas de la distribución de la materia en la región central de las galaxias se deben, de hecho, al análisis de lentes gravitacionales. Este efecto aumenta el brillo aparente y el tamaño angular de las fuentes, facilitando el estudio de objetos que, de lo contrario, serían demasiado débiles para ser detectados. Los objetos responsables de este suceso se denominan lentes o deflectores. El efecto lente producido por lentes extensas, como galaxias y cúmulos de galaxias, que da lugar a imágenes múltiples muy magnificadas y/o distorsionadas de las fuentes se conoce como efecto lente fuerte (*strong lensing*). Los eventos de strong lensing tienen lugar cuando el observador, la lente, y la fuente están suficientemente bien alineados a lo largo de la línea de visión. Para fuentes puntuales, se producirán imágenes múltiples mientras que para emisiones más extendidas, a menudo hay

imágenes distorsionadas, que pueden adquirir la forma de anillos o arcos. La extensión de las fuentes implica que la luz procedente de distantas zonas de la misma se desvía de manera diferente y, por lo tanto, las imágenes producidas se distorsionan. Todas las imágenes pertenecientes a la misma fuente llevan la misma información espectral, y es mediante un análisis espectral como se pueden identificar las imágenes múltiples. El ejemplo más espectacular de imágenes distorsionadas son los arcos gravitacionales observados cerca del centro de muchos cúmulos de galaxias masivos (ver Figuras 0.2 y 3.1).

A lo largo de las últimas décadas, el efecto lente ha demostrado ser una herramienta importante para la Cosmología observacional, dando lugar a un campo de investigación independiente. Los métodos de lente fuerte se basan en el estudio de las magnificaciones, formas, y posiciones de las imágenes multiples de fuentes lensadas, las cuáles se utilizan para acotar la distribución de masa de la lente. Estas imágenes pueden aparecer desde a unos pocos hasta a cientos de kiloparsecs del centro del cúmulo. La interpretación de las observaciones de lentes fuertes en cúmulos de galaxias puede ser difícil debido a las distorsiones introducidas en las galaxias lensadas y a la falta de información (tal como los redshifts de las fuentes, Blandford and Narayan, 1992; Schneider et al., 1992; Wambsganss, 1998; Narayan and Bartelmann, 1996; Kneib, 2002). Afortunadamente, en los últimos años, datos de mayor calidad nos han permitido confirmar espectroscópicamente muchas familias de imágenes múltiples, reduciendo el nivel de incertidumbre y permitiendo reconstrucciones detalladas del efecto lente en muchos cúmulos de galaxias (Broadhurst et al., 2005). Los métodos de reconstrucción del efecto lente se clasifican a grandes rasgos en métodos paramétricos y métodos flexibles (*free-form methods*). Los modelos paramétricos son la elección natural para modelar el efecto lente fuerte en cúmulos cuando el número de observaciones es relativamente pequeño, como en el caso de quasars lensados, donde normalmente solo se tienen tres o cuatro imágenes múltiples. Estos modelos requieren que se hagan suposiciones iniciales sobre la distribución de masa del cúmulo. Algunas de estas suposiciones son, por ejemplo, que los halos de materia oscura siguen a la materia luminosa en el cúmulo, o que los perfiles de las galaxias tienen ciertas simetrías. Sin embargo, si el número de medidas observacionales es suficientemente alto, por ejemplo, de unas pocas decenas a un centenar, es posible reconstruir con precisión el perfil de masa de un cúmulo de galaxias, incluida su subestructura, usando un método flexible. La ventaja de los métodos flexibles es que no se necesitan suposiciones iniciales sobre la distribución de masa en el cúmulo de galaxias. Los modelos del efecto lente derivados con estos métodos son útiles en casos donde la geometría de la lente es compleja, como en el cúmulo MACS0717. En este caso, un modelo de lente flexible pudo predecir correctamente la posición de nuevas familias de galaxias lensadas, que fueron más tarde confirmadas espectroscópicamente (Diego et al., 2015). Incluso en los casos en que el cúmulo es más regular, estos métodos han de-

mostrado ser precisos para predecir diferentes observaciones posteriormente confirmadas. Por ejemplo, predijeron acertadamente el tiempo y la posición de la reaparición de SN Refsdal (Diego et al., 2016), o el redshift de la familia de imágenes lensadas en el sistema 7 del cúmulo A370 (Diego et al., 2018a).

8.5 Análisis del efecto lente fuerte en MACS J1206.2-0847

En esta tesis hemos llevado a cabo un análisis del efecto lente gravitatoria fuerte en el cúmulo de galaxias MACS J1206.2-0847 (J1206 de ahora en adelante) para estimar y acotar la distribución de materia oscura en su región central. J1206 es un cúmulo a $z = 0.439$ muy brillante en rayos X (y por inferencia, muy masivo) en estado dinámico relajado (Biviano et al., 2013). Su BCG es facilmente identificable, y se encuentra en el centro del cúmulo. Una de las imágenes lensadas más peculiares en este cúmulo es un arco gravitacional tangencial muy brillante de unos 15 segundos de arco, extendiendose al oeste de la BCG. El análisis más actual del efecto lente fuerte en este cúmulo ha sido realizado por Caminha et al. (2017) con un enfoque paramétrico flexible. En ese trabajo hicieron uso de 82 imágenes múltiples identificadas espectroscópicamente, pertenecientes a 27 galaxias de fondo con $1.01 \leq z \leq 6.06$.

Hemos usado datos de imágenes públicas obtenidos de los instrumentos ACS y WFC3 del telescopio Hubble. Como miembros del cúmulo hemos seleccionado una muestra de 56 galaxias, incluyendo la BCG. Todas estas galaxias están enumeradas en la Tabla 4.1, y las más cercanas a la BCG se muestran en la Fig. 4.1. Nos hemos limitado a utilizar galaxias brillantes confirmadas espectroscópicamente, cercanas a la BCG y a los arcos observados, que caigan en el campo de visión, y cuyo redshift esté en el intervalo $0.425 \leq z \leq 0.453$. El método de reconstrucción del efecto lente utilizado solo es sensible a las galaxias que están realmente cerca de un arco (unos pocos segundos de arco) o que son muy masivas y, por lo tanto, su radio de influencia efectivo es mayor que unos pocos segundos de arco.

Respecto a las imágenes múltiples observadas, hemos utilizado las mismas identificaciones que en Caminha et al. (2017). Sin embargo, nosotros redefinimos algunos de sus sistemas añadiendo a los arcos alargados bien resueltos diversos marcadores individuales. La adición de marcadores adicionales en sistemas bien resueltos ha demostrado mejorar en gran medida la precisión y estabilidad de las soluciones derivadas por el método WSLAP+ utilzado en este trabajo (Diego et al., 2016) debido a la gran extensión de los arcos gigantes. En concreto, añadimos nuevos marcadores a los arcos 2b (arco tangencial largo situado 20" al oeste de la BCG), 4b (arco radial recto junto a la BCG apuntando al noroeste), y 7c

(arco radial curvo junto a la BCG apuntando al nordeste). En total, hemos usado 97 posiciones de imágenes lensadas en este trabajo. Todas ellas están listadas en la Tabla 4.2, y las más cercanas a la BCG pueden verse en la Fig. 4.2.

El problema fundamental para analizar el efecto lente es averiguar cuáles son las posiciones $\boldsymbol{\beta} = (\beta_x, \beta_y)$ de las galaxias que están siendo lensadas, y la distribución de masa $m(\boldsymbol{\theta})$ del cúmulo, dadas una serie de imágenes lensadas cuyas posiciones $\boldsymbol{\theta} = (\theta_x, \theta_y)$ se conocen. Esto se consigue resolviendo la conocida como ecuación de la lente:

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}, m(\boldsymbol{\theta})) \tag{8.27}$$

donde $\boldsymbol{\alpha}$ es el ángulo de desviación causado por el cúmulo, definido como

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{4G}{c^2}\frac{D_L D_{LS}}{D_S}\int \Sigma(\boldsymbol{\theta'})\frac{\boldsymbol{\theta} - \boldsymbol{\theta'}}{|\boldsymbol{\theta} - \boldsymbol{\theta'}|^2}\, d^2\theta' \tag{8.28}$$

donde $\Sigma(\boldsymbol{\theta})$ es la densidad de masa superficial del cúmulo, y D_L, D_S y D_{LS} son las distancias de diámetro angular al cúmulo, a las galaxias de fondo, y del cúmulo a las galaxias de fondo, respectivamente.

El análisis del efecto lente de J1206 se ha realizado utilizando el código WSLAP+ (Diego et al., 2005a,b, 2007; Sendra et al., 2014). Se trata de un método híbrido puesto que trata Σ como la combinación de dos componentes:

- Una componente difusa, construida como la superposición de N_c funciones Gaussianas distribuidas cada una en una de las celdas de una cuadrícula. En este trabajo se han utilizado una cuadrícula de 1024 celdas iguales (cuadricula regular), y otra cuadrícula con 480 celdas de distinto tamaño (cuadricula adaptativa). Estos dos grids se muestran en la Fig. 4.3.
- Una componente compacta, que da cuenta de la masa asociada con las galaxias individuales en el cúmulo. Para este cúmulo se ha modelado adoptando la distribución de luz observada en el filtro F160W, y asignando inicialmente a cada una de las 56 galaxias seleccionadas una masa proporcional a su brillo superficial. Esta masa es posteriormente reajustada en un proceso de optimización. Esta componente suele dividirse en varias capas, cada una conteniendo un cierto número de galaxias. Esta división permite tratar de diferente manera distintos tipos de galaxias.

Tras descomponer la componente difusa en una serie de funciones de masa individuales, podemos formular la Eq. 8.27 como un sistema de ecuaciones en el que los observables dependen linealmente de todas las incognitas:

$$\boldsymbol{\Theta} = \boldsymbol{\Gamma}\ \boldsymbol{X} \tag{8.29}$$

$$\begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Upsilon}_x & \mathbf{I}_x & \mathbf{0} \\ \boldsymbol{\Upsilon}_y & \mathbf{0} & \mathbf{I}_y \end{pmatrix} \begin{pmatrix} \boldsymbol{c} \\ \beta_x \\ \beta_y \end{pmatrix} \tag{8.30}$$

donde los elementos ij de la matriz $\mathbf{I}_x$ son '1's si el pixel θ_i (coordenada x) procede de la fuente β_j (coordenada y), y cero en los demás casos. La matriz $\mathbf{I}_y$ es definida de forma análoga, y la matriz $\mathbf{0}$ contiene todo ceros. En este sistema de ecuaciones, los observables del efecto lente fuerte (θ_x y θ_y) están contenidos en el array $\boldsymbol{\Theta}$, las posiciones desconocidas de las fuentes (β_x y β_y) y la densidad de masa superficial Σ residen en el array $\boldsymbol{X}$, y la matriz $\boldsymbol{\Gamma}$ da cuenta de la física y la geometría de la cuadrícula utilizada.

La solución a este sistema lineal de ecuaciones es hallada minimizando una función cuadrática de X (ver Eq. 4.14), con la restricción de que dicha solución ha de ser positiva. La convergencia de este algoritmo cuadrático es rápida, permitiendo explorar múltiples soluciones en un tiempo relativamente corto. Pueden obtenerse diferentes soluciones modificando la composición de la componente compacta, la configuración de la cuadrícula y/o los redshifts de los sistemas sin medidas espectroscópicas. Para tener en cuenta las incertidumbres y la variabilidad en los modelos, exploramos una gama de configuraciones en las que cambiamos las suposiciones para los dos componentes principales de nuestro método: la relación luz-masa de las galaxias (componente compacta) y la definición de la cuadrícula (componente difusa). El número de galaxias miembros del cúmulo es siempre 54. Se han considerado 9 tipos de modelos, que se pueden agrupar en cinco categoría o casos, de acuerdo con la definición de la componente compacta:

- Caso 1. Todas las galaxias excepto la BCG están en la misma capa. Se asume que todas las galaxias siguen la misma relación luz-masa. Se han considerado las dos cuadriculas: $N_c = 480$ (1a) y $N_c = 1024$ (1b).
- Caso 2. Todas las galaxias están en la misma capa y siguen la misma relación luz-masa. Se han considerado las dos cuadriculas: $N_c = 480$ (2a) y $N_c = 1024$ (2b).
- Caso 3. Todas las galaxias están en la misma capa y, a excepción de la BCG, siguen una relación luz-masa de 1 a 1. La contribución de la BCG a la masa se reduce a la mitad. Se han considerado las dos cuadriculas: $N_c = 480$ (3a) y $N_c = 1024$ (3b).
- Caso 4. Todas las galaxias están en la misma capa y, a excepción de la BCG, siguen una relación luz-masa de 1 a 1. La contribución de la BCG a la masa se reduce en un tercio. Se han considerado las dos cuadriculas: $N_c = 480$ (4a) y $N_c = 1024$ (4b).

- Caso 5. La BCG es excluida de la componente compacta, y sólo la componente difusa contribuye con su masa a la región central. Las demás galaxias siguen la misma relación luz-masa. Sólo se ha usado la cuadrícula con 480 celdas (5a) puesto que la reconstrucción del arco 7c con la cuadríacula regular no era muy buena.

Todos estos modelos se han comparado entre sí en base a su capadidad para reproducir con mayor precisión dos arcos radiales alargados situados cerca de la BCG: el arco recto 4b (Fig. 4.10) y el arco curvo 7c (Fig. 4.11). El modelo que mejor reproduzca estos dos arcos ofrecerá, en principio, la mejor estimación de la distribución de materia oscura en la región central de J1206. Las métricas utilizadas para comparar los arcos modelados con los datos se han basado en su posición y morfología. Se utilizó un análisis χ^2 para comparar estas métricas. Sorprendentemente, todos los modelos reproducen bastante bien estos dos arcos, independientemente de la configuración elegida para la cuadrícula y la elección de la relación luz-masa de las galaxias. La geometría de estos arcos radiales se debe principalmente a la naturaleza alargada del halo de materia oscura, que es reproducido por todos los modelos. Estimamos, y posteriormente sustraímos, la contribución bariónica estelar para poder estimar la contribución restante de la componente de materia oscura en la región central. Tras comparar las imágenes radiales predichas frente a las observadas para todos los modelos, identificamos el modelo que mejor reproduce la masa en la región central (modelo 2b). En la Fig. 4.14, mostramos una comparativa entre los perfiles de la masa total proyectada de J1206 obtenidos mediante el análisis de la dinámica galáctica en Biviano et al. (2013) (curva roja); el análisis del efecto lente fuerte realizado en Caminha et al. (2017) (curva verde); y en este trabajo (curva rosa), observando una buena compatibilidad entre los tres perfiles. Finalmente, el perfil de densidad de la componente de materia oscura en J1206 se muestra en la Fig. 4.15. Este modelo puede describirse usando un perfil gNFW con radio de escala $r_s = 167$ kpc y pendiente $\gamma_{gNFW} = 0.7$. Nuestra estimación del radio de escala no concuerda con los resultados anteriores de la distribución de masa en este cúmulo, aunque debe tenerse en cuenta que la falta de observaciones a radios altos impide una estimación sólida de este parámetro. Esta estimación conduce a un parámetro de concentración $c_{200} \sim 12$, que cae en el rango predicho para cúmulos relajados. Esto respalda la conclusión de Biviano et al. (2013) de que el cúmulo se encuentra en un estado dinámico relajado. Los resultados de nuestro modelado del efecto lente no son indicativos de un perfil poco plano en la región central del cúmulo, como se esperaría para modelos de materia oscura en los que interactúa consigo misma. Por el contrario, por encima de 10 kpc, encontramos una pendiente cercana a la pendiente esperada para modelos estándar con materia oscura fria. Esta es una conclusión importante ya que nuestro modelo no hace ninguna suposición sobre el perfil de masa en la región interior, aparte de suponer que el componente bariónico traza la masa luminosa.

Usando el modelo 2b hemos sido capaces de predecir la posición de imágenes de fuentes que no fueron utilizadas en nuestro análisis, debido a haber sido observadas cerca de galaxias masivas, y de nuevas imágenes multiples que no han sido observadas (Fig. 4.12). También hemos estimado, usando el modelo 2b, las posiciones de las 27 galaxias siendo lensadas (Fig. 4.16). Entre ellas, podemos destacar los sistemas con ID = 4, 5, 6, y 7. Estas galaxias están todas en un intervalo estrecho de redshift, $1.424 \leq z \leq 1.426$, y probablemente formando un grupo puesto que también están muy cerca posicionalmente. En la Fig. 4.17 se muestra una version reconstruida del plano de la fuente con estas 4 galaxias. Se puede observar cómo las galaxias 4 y 5 están ya interactuando, y que las galaxias 6 y 7 están a menos de 10 kpc de distancia.

8.6 Identificación automática de lentes fuertes

El análisis de los fenómenos de lente gravitatoria fuerte ha demostrado ser una herramienta cosmológica extremadamente útil. El único inconveniente es que estos eventos son poco comúnes, con solo unos pocos sistemas lensados esperados al inspeccionar miles de objetos. El advenimiento de una gran cantidad de datos de la actual y próxima generación de surveys, los cuáles observarán muchos más sistemas lensados, podría parecer que resuelve este problema. Sin embargo, las búsquedas tradicionales en imágenes de las características del efecto lente se han basado exclusivamente en la inspección visual de los objetivos. Y esto no es práctico para tal cantidad de datos. Por lo tanto, el desarrollo de técnicas de búsqueda automática de sistemas lensados ha adquirido una gran importancia en los últimos años. Los métodos de aprendizaje profundo han demostrado tener un gran éxito y pueden utilizarse para una gran variedad de propósitos científicos en astronomía extragaláctica y cosmología. Nosotros hemos encontrado en las redes neuronales convolucionales (CNNs) un buen punto de partida para identificar sistemas fuertemente lensados en imágenes astronómicas.

8.6.1 Redes neuronales convolucionales

Las redes neuronales artificiales (ANNs) son modelos computacionales inspirados en el comportamiento de las redes neuronales biológicas del cerebro humano. Constan de una o más capas de neuronas unidas por conexiones. Estas neuronas son la unidad básica de cálculo en una red neuronal. Cada neurona en una capa recibe uno o más valores de entrada de las neuronas ubicadas en la capa anterior, o de una fuente externa si pertenecen

a la capa de entrada. Cada una de estas entradas tiene un peso asociado (w), que se asigna dependiendo de su importancia relativa respecto a las otras entradas. Todas las neuronas tienen un parámetro de aprendizaje adicional que actúa como compensación de sesgo (b). Las neuronas realizan una combinación lineal ponderada de todos los valores recibidos a la que suman el sesgo. Finalmente, se aplica una función escalar f, conocida como *función de activación*, a esta suma ponderada para calcular una salida escalar y. El proceso descrito se muestra en la Fig. 5.1. En ese esquema, la neurona toma hasta n entradas numéricas diferentes x_i y las asocia con el mismo número de pesos w_i.

Una CNN (Fukushima, 1980; LeCun et al., 1998, 2015) es un tipo de ANN diseñada para aplicarse en imágenes, que esencialmente son matrices de valores en píxeles. Las CNNs son la arquitectura de aprendizaje profundo más popular. Han demostrado ser muy eficaces en los últimos años para tareas de visión artificial, como el reconocimiento de patrones y la clasificación de imágenes (Russakovsky et al., 2015). La principal ventaja de una CNN con respecto a otros algoritmos de reconocimiento de patrones es que detecta y extrae automáticamente las características más representativas de las imágenes sin ninguna supervisión humana. A diferencia de las redes neuronales regulares, la mayoría de las neuronas en las CNNs se organizan en disposiciones tridimensionales (ancho, alto y profundidad). Las neuronas dentro de estas capas 3D solo están conectadas a una pequeña región de la capa anterior. Cada capa transforma los datos de entrada 3D en un volumen de salida 3D. En el boceto que se muestra en la Fig. 5.7, la capa de entrada roja recibe la imagen, por lo que su ancho y alto deben coincidir con las dimensiones de la imagen. Por otro lado, la profundidad será el número de canales de la imagen. Por ejemplo, serían tres (rojo, verde y azul) en el caso de una imagen en color, o mayor en el caso de tener varias imágenes astronómicas de la misma fuente extragaláctica observadas a diferentes frecuencias. La capa de salida final solo tendrá una dimensión, ya que el volumen 3D inicial de imágenes se reduce a un solo vector de puntajes de clase, establecido a lo largo de la dimensión de profundidad. La longitud de este vector dependerá del número de clases posibles en el problema de clasificación.

Además de las capas de entrada y salida, la arquitectura de las CNN consta de otras capas con funciones concretas. La capa convolucional es el núcleo de una CNN, realizando la mayor parte del trabajo computacional. Una CNN típica suele tener más de una capa convolucional. Los parámetros de cada capa convolucional consisten en un conjunto de filtros aprendibles del mismo tamaño y más pequeños que las imágenes de entrada, a lo largo de las dimensiones alto y ancho, pero se extienden en toda su profundidad. Como se puede ver en la Fig. 5.8, cada filtro se desliza sobre las imágenes, calculando, para cada posición del filtro, la suma de la multiplicación entre los elementos del filtro y los elementos de la submatriz de la imagen cubierta por el filtro. El proceso da como resultado

un mapa de activación bidimensional para cada filtro. La red aprenderá filtros capaces de reconocer algún tipo de característica visual presente en las imágenes. Cuantos más filtros tenga la red, más funciones se podrán extraer y mejor será su capacidad de detectar patrones complejos en nuevas imágenes. Después de cada operación de convolución, se aplica una función de activación ReLU. Esta es una operación que reemplaza todos los valores negativos en los mapas de activación por ceros. El propósito es introducir no linealidad en la CNN, ya que la convolución es una operación lineal, pero la mayoría de los datos reales no son lineales.

Es frecuente insertar periódicamente una capa de agrupación después de la capa convolucional. Su función es reducir progresivamente las dimensiones de cada mapa de activación, pero conservando la información más importante. Esto permite disminuir el número de parámetros y trabajo computacional en la red. La capa de agrupación funciona de forma independiente en cada segmento de profundidad de las imágenes y las reduce espacialmente mediante una operación determinada. Existen diferentes tipos de agrupación espacial según el tipo de operación utilizada. El mapa de activación rectificado de entrada se divide en ventanas de las que se toma el elemento más grande (agrupación máxima), el promedio (agrupación promedio) o la suma de todos los elementos (agrupación de suma). Un ejemplo de agrupación máxima se puede observar en la Fig. 5.9.

Las operaciones vistas hasta ahora se encargan de extraer las características útiles de las imágenes, introducir no linealidad en la red y reducir las dimensiones de los mapas de características haciendo que la red sea invariable a pequeñas distorsiones en las imágenes. Finalmente, tras la última capa convolucional, los mapas de características se aplanan como un vector unidimensional y llegan a una o varias capas completamente conectadas. Todas las neuronas de esta capa tienen conexiones con todos los mapas de activación procedentes de las capas anteriores. El objetivo de esta capa es aprender combinaciones no lineales de estas características y clasificar las imágenes de entrada de acuerdo con las clases existentes. Su resultado es una vector de probabilidades para las diferentes etiquetas de clase en que las imágenes pueden ser clasificadas. La clase que recibe la mayor probabilidad se elige como decisión de nuestra clasificación. La Fig. 5.10 muestra un ejemplo de CNN que contiene todas las capas explicadas.

Una vez que se ha decidido la arquitectura de la CNN (número y tipo de capas, número y tamaño de los filtros, etc.) se puede proceder a su entrenamiento. Los demás parámetros y pesos se pueden inicializar con valores aleatorios, o asignarse en base a un conocimiento previo. También tenemos que configurar cómo la red manejará los datos de entrenamiento. Consiste en decidir el tamaño del lote y el número de épocas. El tamaño del lote es el número de porciones en las que se van a dividir los datos de entrenamiento, mientras que la

cantidad de épocas, o ciclos de aprendizaje, es la cantidad de veces que todo el conjunto de datos de entrenamiento pasa por la CNN. El número de iteraciones por época dependerá del tamaño del conjunto de datos de entrenamiento y del tamaño del lote. Por ejemplo, si se tienen 10,000 ejemplos y el tamaño del lote es 50, se necesitarán 200 iteraciones para completar 1 época. En cada época cambia el orden y la composición de los lotes, por lo que la red ve los ejemplos en un orden diferente en cada ciclo. Una vez decidido esto, las imágenes de entrenamiento se pueden introducir en la red, pasando por todas sus capas, hasta encontrar para cada imagen de entrada las probabilidades de pertenecer a cada clase posible. El error total sobre todas las clases en la capa de salida se evalúa utilizando una función de pérdida. Un método de retropropagación se usa para calcular los gradientes de la función de pérdida con respecto a todos los pesos en la red, y un algoritmo de descenso de gradiente actualiza todos los pesos para minimizar el error en la capa de salida, como se muestra en la Ec. (5.4). Una vez que se alcanzan los pesos óptimos, podemos decir que la CNN ha aprendido a clasificar correctamente las imágenes del conjunto de datos de entrenamiento y está lista para trabajar con un conjunto de datos de prueba.

Sin embargo, a veces el aprendizaje no es exitoso y la red ni siquiera clasifica lo suficientemente bien la muestra de entrenamiento, o funciona mucho mejor con los datos de entrenamiento que con los de prueba. En este último caso, se dice que el modelo está sobreajustado. Este tipo de problemas de aprendizaje se pueden diagnosticar fácilmente monitorizando el desempeño de la CNN a lo largo del tiempo durante el entrenamiento. Entre las técnicas existentes para evitar el sobreajuste, el abandono es, con mucho, la técnica de regularización más popular. Como se muestra en la Fig. 5.6, el abandono consiste en asignar a todas o algunas de las neuronas de la red una cierta probabilidad de quedar temporalmente desactivadas durante el tiempo de entrenamiento, de modo que la arquitectura de la red cambiará en cada iteración. La otra forma habitual de reducir el sobreajuste es aumentar el número de ejemplos en los datos de entrenamiento mediante el 'aumento de datos'. Esto consiste en aplicar a las imagenes diferentes técnicas como inversiones, traslaciones, rotaciones, escalados, cambios en el brillo o la inclusión de ruido.

El rendimiento de cualquier modelo de clasificación, como los producidos por las redes neuronales, sobre unos datos de prueba, de los que se conocen los resultados verdaderos, puede describirse usando una matriz de confusión. Esta tabla muestra el número de aciertos y de fallos para cada clase, así como qué clases reciben los fallos del resto. La suma de todas las celdas en la matriz de confusion debe coincidir con la cantidad de ejemplos en la muestra de prueba. En el caso concreto de una clasificación binaria, como la que consideramos aquí para intentar predecir si una imágen es o no una lente, se establece un valor de umbral para decidir si un ejemplo pertenece a una u otra clase. El valor por defecto para dicho umbral, sobre el que son construidas las matrices de confusión, es

0.5. Por tanto, los ejemplos con probabilidades mayores que 0.5 son predichos como lentes mientrás que los demás se consideran no lentes. Todos los ejemplos de los datos que son lentes se denominan positivos (P) mientrás que los que no lo son se denominan negativos (N). Los verdaderos positivos (TP) son los ejemplos de lentes identificados correctamente como lentes. Los verdaderos negativos (TN) son los ejemplos de no lentes identificados correctamente como no lentes. Los falsos positivos (FP) son los ejemplos de no lentes que la red predice erróneamente como lentes. Los falsos negativos (FN) son los ejemplos de lentes que la red predice erróneamente como no lentes. Un ejemplo de las matrices de confusión utilizadas en este trabajo es mostrado en la Tabla 5.1.

Estamos ya en disposición de definir algunas tasas importantes que se pueden calcular a partir de la información dada por la matriz de confusión. La tasa de verdaderos positivos (TPR o $Recall$) indica la probabilidad de detección y se define como el cociente TP/P (Eq. 5.12). La tasa de falsos positivos (FPR) nos informa de la probabilidad de una falsa alarma, y se define como el cociente FP/N (Eq. 5.13). La exactitud (Acc) indica cuántas de las predicciones hechas por el clasificador son correctas (Eq. 5.14). Y por su lado tenemos la precisión (Pre), que nos dice la proporción de lentes correctamente identificadas con respecto al total de identificaciones de lentes realizadas sobre los datos (Eq. 5.15). Es, por tanto, un indicador de la pureza de la clasificación realizada.

El método más comúnmente utilizado para cuantificar la habilidad clasificadora de un modelo es la curva ROC (Característica Operativa del Receptor; Dorfman and Alf, 1968; Powers, 2011). Esta curva representa FPR frente a TPR para diferentes umbrales de probabilidad (P_{thr}). Algunos ejemplos de curvas ROC se pueden ver en la Fig. 6.4. El AUC (área bajo la curva) proporcionada por esta curva es el factor de mérito estándar utilizado para comparar el desempeño de varios modelos de clasificación. Indica cuánto es capaz un modelo de distinguir entre clases. Buscamos maximizar los TP y minimizar los FP. Por lo tanto, cuanto más hacia la izquierda en el eje x y hacia arriba en el eje y esté la curva, mayor será el AUC y mejor será el modelo.

Finalmente, una vez introducidas las cantidades anteriores, podemos definir F_β:

$$F_\beta = (1+\beta^2)\frac{Pre \times Recall}{(\beta^2 Pre + Recall)} \tag{8.31}$$

de modo que $F_{\beta=0} = Pre$, $F_{\beta=\infty} = Recall$, y $0 \leq F_\beta \leq 1$. $F_\beta = 0$ cuando $TP = 0$, mientrás que $F_\beta = 1$ si no hay FP ni FN. La importancia relativa en esta expresión de Pre y $Recall$ depende del valor que tome β. F_β puede calcularse para un valor P_{thr} específico, y el funcionamiento del clasificador se mide usando el valor máximo de F_β alcanzado para cualquier P_{thr}:

$$F_\beta = \max_p F_\beta(p) \tag{8.32}$$

Cuando se buscan sistemas lensados, tener una alta precisión, o una baja tasa de contaminación, es importante debido al hecho de que en los datos reales las lentes son más raras de ver que en los datos simulados por un factor de ~ 1000. Es por ello que en este trabajo hemos tomado $\beta = 0.03$ para priorizar *Pre* sobre el *Recall*.

La configuración de CNN utilizada en este trabajo es una adaptación de la red desarrollada para la clasificación morfológica de galaxias en Domínguez Sánchez et al. (2018), ahora dirigida a la identificación de lentes fuertes. Ha sido construida y entrenada en Python 3.6 usando la biblioteca Keras, una interfaz de programación de alto nivel y código abierto para la aplicación de redes neuronales. La arquitectura básica de esta CNN se muestra en la Fig. 5.11. Se compone de 4 capas convolucionales con filtros cuadrados de diferentes tamaños (6×6, 5×5, 2×2 y 3×3, respectivamente) y una capa completamente conectada. Después de cada capa convolucional, se aplica una función de activación ReLu. Y después de la segunda y tercera capas convolucionales, se aplica una agrupación máxima de 2×2. Al principio, el abandono se realizaba después de cada capa convolucional con tasas de 0.25 o 0.5, como se muestra en el esquema. Sin embargo, después de realizar algunas pruebas con simulaciones del efecto lente similares a los que observaría el survey KiDS, decidimos eliminar todos los abandonos excepto el de la capa completamente conectada. La arquitectura de la red se mantuvo así desde ese momento.

Hemos utilizado diferentes datos de entrada para la red a lo largo de este trabajo y, en consecuencia, la capa de entrada no siempre ha sido la misma. La capa de entrada que se muestra en la Fig. 5.11 corresponde a la red utilizada en la Sec. 6.1, donde hemos entrenado usando imágenes similares a KiDS, de 101×101 píxeles, en tres bandas diferentes (g, r y i). Por tanto, las matrices leídas por la red tienen dimensiones (101,101,3). En la Sec. 6.2 entrenamos usando imágenes similares a las que observará el satélite Euclid, también de 101×101 píxeles, solo en la banda visible, por lo que las matrices leídas por la red tienen dimensiones (101,101,1). En Sec. 6.3 entrenamos la red nuevamente usando simulaciones inspiradas en Euclid, pero con 66×66 píxeles, y la imagen está disponible en cuatro bandas diferentes (VIS, J, Y y H). Las matrices leídas por la red esta vez tienen dimensiones (66,66,4). Finalmente, en la Sec. 7.4 la red se entrena usando dos imágenes pseudo-espectrales, con 56×56 píxeles, para cada ejemplo por clasificar. Por tanto, las matrices a introducir en la red tienen dimensiones (56,56,2). En todos los casos, los valores de flujo de las imágenes se normalizan al máximo en cada banda antes de ingresar en la red, para no trabajar con números grandes.

Los modelos se entrenan utilizando una tabla con los resultados esperados (aprendizaje supervisado) como clasificadores binarios, entre lentes (1) y no lentes (0). Se usa una función de activación sigmoide en la última capa completamente conectada para convertir los puntajes de salida en probabilidades distribuidas entre 0 y 1. En el proceso de aprendizaje usamos la función de pérdida de entropía cruzada binaria (Eq. 5.18). El método de descenso del gradiente utilizado aquí es la optimización adam (Kingma and Ba, 2015) con una tasa de aprendizaje de 0.001. El resultado una vez que se ha entrenado la red es un vector de puntuaciones de probabilidad entre 0 y 1, cuya longitud concuerda con el número de ejemplos dados a la red. Una probabilidad de 0 significa la menor confianza de que el ejemplo dado es una lente, mientras que una probabilidad de 1 significa la confianza más alta de que se trata de una lente. Usamos 25 épocas de entrenamiento al probar diferentes estrategias para trabajar con la red, como cambiar el tamaño de las muestras de entrenamiento o probar diferentes preprocesados de datos. Y 60 épocas, con la opción de detener el proceso de forma anticipada si la optimización ha convergido, cuando se lleva a cabo el entrenamiento definitivo para un determinado conjunto de datos. El tamaño del lote se fija siempre en 30. En el proceso de entrenamiento, se realizan varias técnicas de aumento de datos, acercando y alejando las imágenes (0.75 a 1.3 veces el tamaño original), rotándolas (dentro de 45°), volteándolas y desplazándolas tanto vertical como horizontalmente (en un 5%). Esto asegura que los modelos derivados no sufran de sobreajuste ya que los datos de entrada varían de una época de entrenamiento a la siguiente. La proporción de lentes/no lentes no está necesariamente equilibrada en los conjuntos de datos de entrenamiento utilizados. Teniendo esto en cuenta, nos aseguramos de equilibrar los pesos asignados a las clases de lentes y no lentes durante el entrenamiento para evitar sesgos no deseados en el desempeño de la CNN.

8.6.2 Buscando lentes fuertes con redes neuronales convolucionales

En la última parte de esta tesis se ha presentado el desempeño en la búsqueda de lentes gravitacionales fuertes de varios modelos basados en la misma red neuronal convolucional (Fig. 5.11). Hemos considerado, tanto en el entrenamiento como en el testeo, datos que imitan la respuesta o proceden de diferentes surveys astronómicos. En primer lugar, se llevaron a cabo varias pruebas con datos simulados basados en KiDS (de Jong et al., 2013). Gracias a estas pruebas descubrimos que se obtiene un mejor rendimiento de la red al normalizar previamente los datos al valor máximo en cada una de las bandas consideradas, reducir la frecuencia de la técnica de abandono, y usar la técnica de detención anticipada cuando no se observan mejoras en la muestra de validación. Aproximadamente el 13% de

los ejemplos basados en KiDS presenta un enmascaramiento artificial aleatorio en algunos píxeles, que consideramos se incluyeró para imitar un excesivo ruido en los datos o una mala respuesta del detector. No obstante, al considerar estas máscaras excesivas, puesto que ocultan en bastantes casos por completo las características de las lentes presentes en las imágenes, y al lograr mejores resultados entrenando sin ellas, decidimos eliminar estos ejemplos de los datos de entrenamiento, y crear un muestra de datos de prueba sin ellos.

Nuestro mejor modelo obtenido entrenando con datos basados en KiDS ha logrado un AUC = 0.955 cuando se prueba en el conjunto de datos de prueba completo, y un AUC = 0.982 cuando se eliminan los ejemplos que contienen máscaras artificiales (ver Fig. 6.16). Un 15% de los ejemplos simulados en estos datos de prueba utilizan imágenes reales tomadas de una muestra preliminar de galaxias brillantes observadas con el survey KiDS. Cuando nuestro modelo se aplica a estos ejemplos concretos, el AUC cae a 0.832, pero con un modelo entrenado y probado utilizando exclusivamente estos ejemplos logramos un AUC = 0.888 (ver Fig. 6.17). Este es un claro ejemplo de las diferencias de eficiencia que pueden surgir cuando los modelos entrenados mediante simulaciones son puestos a prueba con datos reales. Y, por tanto, es importante que las simulaciones sean lo más realistas posible. Posteriormente, hemos probado nuestro mejor modelo en imágenes de una muestra de candidatos a lentes gravitacionales fuertes de los surveys SDSS y DES, logrando precisiones del 65% y 83%, respectivamente. En las figuras 6.18, 6.19 y 6.20 se muestran imagenes de algunos de estos sistemas junto a las probabilidades asignadas por nuestro modelo. Hay ejemplos tanto bien como mal clasificados. Por último, este modelo también se ha aplicado a los 40 posibles candidatos a lentes identificados en la Sección 2.7, entre los que se pronostican 7 como lentes (ver en Fig. C aquellos ejemplos con el nombre resaltado en naranja), y a un conjunto de posibles lentes identificadas visualmente gracias al proyecto Galaxy Zoo 2 (Willett et al., 2013). El 86% de estos últimos ejemplos han sido identificados como lentes por nuestro modelo (ver Fig. 6.21).

Por otro lado, el mejor modelo obtenido entrenando con datos similares a Euclid en la banda visible arroja un AUC = 0.921 (ver Fig. 6.25). Este modelo se obtuvo usando criterios ligeramente más estrictos para definir las lentes que los utilizados para evaluar dicho modelo en el conjunto de datos de prueba. Como puede verse, a pesar del hecho de que las imágenes basadas en Euclid tienen una relación S/R más alta que las imágenes basadas en KiDS, los resultados del AUC son mucho mejores para estas últimas. Después de entrenar un modelo solo con imágenes en la banda r de los datos basados en KiDS, podemos confirmar que proporcionar a la red más información, en forma de bandas adicionales, mejora su rendimiento. Por lo tanto, la razón por la que la CNN funciona peor con los datos de Euclid es que solo poseen una banda.

El siguiente paso ha sido entrenar y probar la red con datos basados en Euclid más realistas, utilizando sus cuatro bandas previstas (VIS, J, Y y H), como parte un desafio entre diversos métodos dirigidos a la identificación de sistemas lensados. En esta ocasión nos propusimos lograr un modelo con una alta precisión, ya que el método de evaluación utilizado en el desafío penalizaba la cantidad de falsos positivos. Se consideraron varios modelos, que fueron entrenados con definiciones más agresivas para las lentes que la utilizada en la evaluación, entre los cuales el '50-pixel model' proporcionó los mejores resultados con $F_\beta = 0.9859$ (ver Fig. 6.35) y AUC $= 0.700$ (ver Fig. 6.34). Logramos obtener una clasificación de lentes lo suficientemente pura en el desafío, pero a costa de perder más lentes de las detectadas (ver Tabla 6.24). Mostramos en la Fig. 6.37 una muestra aleatoria de 35 ejemplos de verdaderos positivos con $P_{lens} = 1.0$. Después de comprobar visualmente algunos falsos positivos con probabilidades altas, encontramos que $\approx 33\%$ son verdaderas lentes reconocibles a simple vista, pero que no cumplen con los requisitos de evaluación del desafío (ver Fig. 6.38). Estas son noticias alentadoras, ya que nuestros modelos agresivos son capaces de reconocer patrones de sistemas lensados en ejemplos menos claros que los utilizados en su entrenamiento.

Tenemos la intención de seguir explorando la capacidad de los algoritmos basados en CNNs. Nuestro objetivo a largo plazo es desarrollar modelos buscadores de lentes automatizados lo más precisos posible, que sean capaces de extraer características difíciles de reconocer para el ojo humano. Otro aspecto importante es asegurar que estos modelos no dependan del instrumento ni de las características de la encuesta utilizadas para obtener las imágenes. Hemos visto aquí que es posible aplicar modelos entrenados con datos basados en KiDS a imágenes basadas en Euclid, o procedentes de SDSS y DES, logrando buenos resultados. Los modelos derivados deben ser robustos frente a pequeños cambios en los datos. Abordaremos esto explorando todas las posibilidades de las CNN y combinando las lentes ya detectadas de diferentes surveys con imágenes simuladas de sistemas fuertemente lensados en diferentes condiciones.

Finalmente, hemos mostrado los primeros resultados de un trabajo en curso destinado a desarrollar un método de identificación automática de cuásares lensados en J-PAS (Bonoli et al., 2020). J-PAS es un survey único con 54 filtros fotométricos estrechos que nos permitirá tener un buen conocimiento de la SED de los objetos observados, así como medidas fotométricas precisas de sus redshifts. Se espera que explore unos 8000 grados cuadrados y observe, entre otros objetos, más de medio millón de cuásares, entre los que una cantidad no despreciable estará afectada por el efecto lente gravitatoria. En este trabajo hemos creado simulaciones del efecto lente gravitatoria sobre cuásares, también simulados, basados en el sistema fotométrico de J-PAS. Como lentes se han usado galaxias rojas detectadas en un pequeño campo de visión de J-PAS ya explorado (ver Fig. 7.6). Estas sim-

ulaciones del efecto lente, junto con ejemplos no lensados de galaxias, quásares y estrellas observados con J-PAS, se han utilizado para entrenar nuestra CNN. En lugar de entrenar la red con las imágenes en todas las bandas consideradas, reducimos toda esta información a dos pseudo-espectros para cada ejemplo. Estos pseudo-spectros fueron construidos proyectando los datos en todas las bandas a lo largo de los ejes x e y de las imágenes (ver Fig. 7.5). Aunque se pierda parte de la información, las características morfológicas y espectrales más relevantes de los ejemplos, para poder clasificarlos, se conservan, y se reduce la complejidad de la red. En la Fig. 7.7 se muestran ejemplos de los pseudo-spectros de algunos cuásares lensados. Y en las figuras 7.8 y 7.9 se muestran los pseudo-spectros de algunos ejemplos de estrellas/cuásares y galaxias, respectivamente. Hemos comparado el desempeño de dos modelos, uno entrenado con galaxias, quásares y estrellas, como ejemplos sin lentes, y el otro entrenado sólo con galaxias, como casos no lensados. Encontramos que ambos modelos logran resultados muy buenos (ver Fig. 7.11), pero observamos que entrenar con una mayor diversidad de ejemplos reduce significativamente la cantidad de falsos positivos (ver Tabla 7.5) y la probabilidad que el modelo asigna a los mismos (ver Tabla 7.6).

Nuestro deseo es comprobar a continuación el rendimiento de la CNN al aumentar la cantidad de cuásares lensados con magnificaciones pequeñas, que seguramente serán menos fáciles de clasificar. Aumentar el número de estrellas, cuásares y galaxias considerados como ejemplos no lensados también está entre nuestros propósitos. A más largo plazo, nuestro objetivo es probar los modelos que obtengamos sobre las fuentes detectadas en el campo de visión completo de J-PAS. Las identificaciones correctas logradas de cuásares lensados serán de interés para el survey WEAVE-QSO (Pieri et al., 2016), cuyos objetivos serán los cuásares observados por J-PAS.

8.7 Conclusiones y proyección futura

No ha transcurrido mucho tiempo desde que las redes neuronales convolucionales comenzaron a usarse en cosmología, pero han logrado convertirse en una de las herramientas más atractivas para analizar datos, en particular para fines de clasificación. Y seguramente su relevancia seguirá creciendo cada vez más debido a su capacidad para detectar características sin supervisión humana. Hemos mostrado en este trabajo cómo pueden emplearse con éxito para buscar lentes fuertes en imágenes astronómicas. Pero también son utilizadas para distinguir entre diferentes fuentes astronómicas o radiaciones, en la clasificación morfológica de galaxias o para medir redshifts fotométricos. Para todos estos propósitos, tener un buen conocimiento de la SED observada es crucial para que la red pueda caracterizar

la naturaleza o estimar el redshift de los objetos detectados. Aunque los surveys espectroscópicos permiten análisis precisos de las fuentes, se requieren prolongadas campañas de seguimiento para cada objetivo. Por otro lado, los grandes surveys fotométricos proporcionan datos para tantas fuentes como aparezcan en las imágenes, pero la resolución en longitud de onda depende de cuán estrechos sean los filtros, perdiéndose información sobre las características espectrales individuales. Por esta razón, es gratificante la existencia de surveys con bandas estrechas como J-PAS o PAU (Eriksen et al., 2019). Con sus 54 filtros estrechos, J-PAS permite aumentar la precisión produciendo un espectro completo de baja resolución, que cubre una amplia gama de longitudes de onda visibles. Si bien las técnicas automáticas de aprendizaje profundo, como las redes neuronales convolucionales, nos ayudan a manejar más fácilmente las grandes cantidades de datos procedentes de los surveys a gran escala actuales y futuros, los pseudo-espectros de J-PAS son útiles para resolver las amplias características espectrales de la objetos observados, más rápido que los obtenidos con surveys espectroscópicos. Esperamos que se lleven a cabo numerosos estudios astrofísicos mediante la combinación de estos dos ingredientes, produciendo significativos resultados científicos. Aunque las redes neuronales son realmente poderosas por sí mismas, hemos visto que un preprocesamiento adecuado de los datos es importante. En este sentido, la combinación de técnicas adaptativas de filtrado concretas, como el multifiltro ajustado, que optimiza la detección de fuentes sobre ruido de fondo, y redes neuronales convolucionales también sería un proyecto interesante a desarrollar en el futuro.

Centrándonos en el efecto lente gravitatoria, es un fenómeno que involucra dos objetos separados en redshift y con diferentes comportamientos espectrales, en los que generalmente se pueden ver estructuras azules tenues en forma de arco alrededor de galaxias rojas luminosas. En este trabajo hemos confirmado que tener observaciones a diferentes longitudes de onda arroja mejores resultados al buscar lentes fuertes. Por lo tanto, la gran cantidad de bandas fotométricas de J-PAS puede proporcionar una visión clara para detectar sistemas lensados, distinguir la fuente de la lente debido a sus diferencias espectrales e identificar múltiples imágenes de la misma fuente debido a la similitud espectral en las imágenes observadas. Uno de los principales objetivos de J-PAS es medir redshifts fotométricos precisos para galaxias de hasta $z \sim 1$, lo que facilita significativamente el descubrimiento de sistemas lensados para que puedan ser estudiados posteriormente por este u otros surveys. Cuantas más observaciones estén disponibles, mejores análisis del efecto lente se pueden realizar. Y esto es importante, ya que las estadísticas del efecto lente son una forma prometedora de derivar restricciones sobre los parámetros cosmológicos. Gracias a las técnicas de detección automática, como las redes neuronales convolucionales, y a los surveys a gran escala, como Euclid y LSST, que esperan observar $\sim 10^5$ lentes fuertes, la búsqueda y análisis de eventos con lentes fuertes es un campo de investigación prometedor.

Bibliography

H. M. Abdelsalam, P. Saha, and L. L. R. Williams. Non-parametric reconstruction of cluster mass distribution from strong lensing: modelling Abell 370. *MNRAS*, 294:734–746, 1998a. doi:10.1046/j.1365-8711.1998.01356.x.

H. M. Abdelsalam, P. Saha, and L. L. R. Williams. Non-parametric reconstruction of Abell 2218 from combined Weak and Strong Lensing. *AJ*, 116(4):1541–1552, 1998b. doi:10.1086/300546.

B. Abolfathi, D. S. Aguado, G. Aguilar, C. Allende, A. Almeida, T. Tasnim, and F. Anders. The Fourteenth Data Release of the Sloan Digital Sky Survey. *ApJS*, 235:42, 2018. doi:10.3847/1538-4365/aa9e8a.

R. Akeson et al. The Wide Field Infrared Survey Telescope, 2019. arXiv:1902.05569.

C. Alard. Automated detection of gravitational arcs. June 2006. arXiv:astro-ph/0606757v1.

K. E. Andrade, Q. Minor, A. Nierenberg, and M. Kaplinghat. Detecting dark matter cores in galaxy clusters with strong lensing. *MNRAS*, 487:1905–1926, 2019. doi:10.1093/mnras/stz1360.

J. S. Arabadjis, M. W. Bautz, and G. P. Garmire. Chandra Observations of the Lensing Cluster EMSS 1358+6245: Implications for Self-interacting Dark Matter. *ApJ*, 572(1): 66–78, June 2002. doi:10.1086/340296.

T. Bakx et al. The Herschel Bright Sources (HerBS): Sample definition and SCUBA-2 observations. *MNRAS*, 473:1751, 2018. doi:10.1093/mnras/stx2267.

V. E. Barnard, P. Vielva, et al. The very bright SCUBA galaxy count: looking for SCUBA galaxies with the Mexican hat wavelet. *MNRAS*, 352:961–974, Aug. 2004. doi:10.1111/j.1365-2966.2004.07985.x.

R. B. Barreiro, J. L. Sanz, D. Herranz, and E. Martínez-González. Comparing filters for the detection of point sources. *MNRAS*, 342:119, 2003. doi:10.1046/j.1365-8711.2003.06520.x.

M. Bartelmann. Arcs from a universal dark-matter halo profile. *A&A*, 313:697, 1996. arXiv:astro-ph/9602053v1.

M. Bartelmann and P. Schneider. Weak gravitational lensing. *Physics Reports*, 340(4-5): 291–472, Jan 2001. doi:10.1016/s0370-1573(00)00082-x.

M. Bartelmann, A. Huss, J. M. Colberg, A. Jenkins, and F. R. Pearce. Arc statistics with realistic cluster potentials. IV. Clusters in different cosmologies. *A&A*, 330:1–9, Feb. 1998. arXiv:astro-ph/9707167.

J. P. Beaulieu et al. Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing. *Nature*, 439(7075):437–440, Jan. 2006. doi:10.1038/nature04441.

P. Bergamini, P. Rosati, et al. Enhanced cluster lensing models with measured galaxy kinematics. *A&A*, 631:16, 2019. doi:10.1051/0004-6361/201935974.

F. Bianchini, A. Lapi, M. Calabrese, P. Bielewicz, J. Gonzalez-Nuevo, et al. Toward a tomographic analysis of the cross-correlation between Planck CMB lensing and H-ATLAS galaxies. *ApJ*, 825:24, 2016. doi:10.3847/0004-637X/825/1/24.

F. Bianchini, G. Fabbian, A. Lapi, J. Gonzalez-Nuevo, R. Gilli, and C. Baccigalupi. Broadband spectral energy distributions of SDSS-selected quasars and of their host galaxies. *ApJ*, 871, June 2018. doi:10.3847/1538-4357/aaf86b.

A. D. Biggs, I. W. A. Browne, P. Helbig, L. V. E. Koopmans, P. N. Wilkinson, and R. A. Perley. Time delay for the gravitational lens system B0218+357. *MNRAS*, 304(2): 349–358, 04 1999. doi:10.1046/j.1365-8711.1999.02309.x.

J. Binney and S. Tremaine. *Galactic Dynamics*. 2008. doi:10.1515/9781400828722.

A. Biviano, P. Rosati, I. Balestra, A. Mercurio, M. Girardi, et al. CLASH-VLT: The mass, velocity-anisotropy, and pseudo-phase-space density profiles of the $z = 0.44$ galaxy cluster MACS J1206.2–0847. *A&A*, 558:A1, 2013. doi:10.1051/0004-6361/201321955.

A. W. Blain and M. S. Longair. Submillimetre Cosmology. *MNRAS*, 264:509, Sept. 1993. doi:10.1093/mnras/264.2.509.

R. D. Blandford and C. S. Kochanek. Gravitational Imaging by Isolated Elliptical Potential Wells. I. Cross Sections. *ApJ*, 321:658, Oct. 1987. doi:10.1086/165660.

R. D. Blandford and R. Narayan. Cosmological applications of gravitational lensing. *Annu. Rev. Astron. Astrophys.*, 30:311, 1992. doi:10.1146/annurev.aa.30.090192.001523.

M. R. Blanton and J. Moustakas. Physical Properties and Environments of Nearby Galaxies. *Annu. Rev. Astron. Astrophys.*, 47(1):159–210, Sept. 2009. doi:10.1146/annurev-astro-082708-101734.

M. R. Blanton et al. Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe. *AJ*, 154:28, 2017. doi:10.3847/1538-3881/aa7567.

H. Böhringer et al. The ROSAT-ESO flux limited x-Ray (REFLEX) galaxy cluster survey I: The construction of the cluster sample. *A&A*, 369:826, 2001. doi:10.1051/0004-6361:20010240.

H. Böhringer et al. The ROSAT-ESO flux limited X-ray (REFLEX) galaxy cluster survey V: The cluster catalogue. *A&A*, 425:367–383, 2004. doi:10.1051/0004-6361:20034484.

A. S. Bolton et al. The Sloan Lens ACS Survey. V. The Full ACS Strong–Lens Sample. *ApJ*, 682(2):964–984, Aug 2008. doi:10.1086/589327.

S. Bonoli et al. The miniJPAS survey: a preview of the Universe in 56 colours. 2020. arXiv:2007.01910.

N. Bourne et al. Colour matters: the effects of lensing on the positional offsets between optical and submillimetre galaxies in Herschel-ATLAS. *MNRAS*, 444:2, 2014. doi:10.1093/mnras/stu1582.

N. Bourne et al. The Herschel-ATLAS Data Release 1 Paper II. Multi-wavelength counterparts to submillimetre sources. *MNRAS*, 462:1714–1734, 2016. doi:10.1093/mnras/stw1654.

M. Bradač et al. Strong and weak lensing united. II. The cluster mass distribution of the most X-ray luminous cluster RX J1347.5-1145. *A&A*, 437(1):49–60, July 2005. doi:10.1051/0004-6361:20042234.

T. Broadhurst et al. Strong-Lensing Analysis of A1689 from Deep Advanced Camera Images. *ApJ*, 621(1):53–88, March 2005. doi:10.1086/426494.

T. J. Broadhurst, A. N. Taylor, and J. A. Peacock. Mapping Cluster Mass Distributions via Gravitational Lensing of Background Galaxies. *ApJ*, 438:49, Jan. 1995. doi:10.1086/175053.

G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. *MNRAS*, 344:1000–1028, 2003. doi:10.1046/j.1365-8711.2003.06897.x.

R. S. Bussmann, M. A. Gurwell, H. Fu, D. J. B. Smith, S. Dye, R. Auld, and M. Baes. A Detailed Gravitational Lens Model Based on Submillimeter Array and Keck Adaptive Optics Imaging of a Herschel-ATLAS Sub-millimeter Galaxy at z=4.243. *ApJ*, 756:134, 2012. doi:10.1088/0004-637X/756/2/134.

R. S. Bussmann, I. Perez-Fournon, S. Amber, J. A. Calanog, M. A. Gurwell, H. Dannerbauer, and F. De Bernardis. Gravitational Lens Models Based on Submillimeter Array Imaging of Herschel-selected Strongly Lensed Sub-millimeter Galaxies at z>1.5. *ApJ*, 779:25, 2013. doi:10.1088/0004-637X/779/1/25.

R. A. Cabanac et al. The CFHTLS strong lensing legacy survey. I. Survey overview and T0002 release sample. *A&A*, 461(3):813–821, Jan. 2007. doi:10.1051/0004-6361:20065810.

G. B. Caminha, C. Grillo, P. Rosati, M. Meneghetti, A. Mercurio, et al. Mass distribution in the core of MACS J1206: robust modeling from an exceptionally large sample of central multiple images. *A&A*, 607:A93, 2017. doi:10.1051/0004-6361/201731498.

J. A. Canalog et al. Lens Models of Herschel-Selected Galaxies from High-Resolution Near-IR Observations. *ApJ*, 797:138, 2014. doi:10.1088/0004-637X/797/2/138.

J. E. Carlstrom, P. A. R. Ade, K. A. Aird, B. A. Benson, L. E. Bleem, S. Busetti, et al. The 10 Meter South Pole Telescope. *PASP*, 123(903):568, May 2011. doi:10.1086/659879.

A. J. Cenarro et al. The Observatorio Astrofísico de Javalambre: Current status, developments, operations, and strategies. In *Observatory Operations: Strategies, Processes, and Systems V*, volume 9149 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 91491I, Aug. 2014. doi:10.1117/12.2055455.

A. J. Cenarro et al. Commissioning and first scientific operations of the wide-field 2.6m Javalambre Survey Telescope. In *Ground-based and Airborne Telescopes VII*, volume 10700, pages 93 – 109. International Society for Optics and Photonics, SPIE, 2018. doi:10.1117/12.2309520.

G. Chabrier. Galactic Stellar and Substellar Initial Mass Function. *PASP*, 115:763–795, 2003. doi:10.1086/376392.

A. Challinor et al. Exploring cosmic origins with CORE: Gravitational lensing of the CMB. *JCAP*, 2018(04):018–018, Apr 2018. doi:10.1088/1475-7516/2018/04/018.

E. L. Chapin et al. A joint analysis of BLAST 250-500 μm and LABOCA 870 μm observations in the Extended Chandra Deep Field-South. *MNRAS*, 411:505–549, Feb. 2011. doi:10.1111/j.1365-2966.2010.17697.x.

M. Chiba and T. Futamase. Statistics of QSO Lensing and Cosmological Parameters. *PROG*, 133:115–135, Jan. 1999. doi:10.1143/PTPS.133.115.

A. Choromanska, M. Henaff, M. Mathieu, G. Ben Arous, and Y. LeCun. The Loss Surfaces of Multilayer Networks. *JMLR*, 12:192, 2015. arXiv:1412.0233v3.

D. Ciresan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In *In Proceedings of the 25th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012)*, pages 3642–3649, 2012. doi:10.1109/CVPR.2012.6248110.

D. L. Clements et al. H-ATLAS: A candidate high redshift cluster/protocluster of star-forming galaxies. *MNRAS*, 461(2):1719–1733, Sept. 2016. doi:10.1093/mnras/stw1224.

D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky. A Direct Empirical Proof of the Existence of Dark Matter. *ApJ*, 648(2): L109–L113, Aug 2006. doi:10.1086/508162.

T. E. Collett. The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys. *ApJ*, 811(1):20, Sept. 2015. doi:10.1088/0004-637X/811/1/20.

A. Conley, A. Cooray, J. D. Vieira, E. A. González-Solares, J. E. Aguirre, and A. Amblard. Discovery of a Multiply-Lensed Submillimeter Galaxy in Early HerMES Herschel/SPIRE Data. *ApJL*, 732:35, 2011. doi:10.1088/2041-8205/732/2/L35.

P. Cox, M. Krips, R. Neri, A. Omont, R. Gusten, K. M. Menten, F. Wyrowski, A. Weiss, and A. Beelen. Gas and dust in a submillimeter galaxy at $z = 4.24$ from the Herschel ATLAS. *ApJ*, 740:63, 2011. doi:10.1088/0004-637X/740/2/63.

A. Davies, S. Serjeant, and J. M. Bromley. Using convolutional neural networks to identify gravitational lenses in astronomical images. *MNRAS*, 487(4):5263–5271, Aug. 2019. doi:10.1093/mnras/stz1288.

J. T. A. de Jong, G. A. Verdoes Kleijn, K. H. Kuijken, and E. A. Valentijn. The Kilo-Degree Survey. *Exp. Astron*, 35(1-2):25–44, Jan. 2013. doi:10.1007/s10686-012-9306-1.

G. de Zotti, M. Massardi, M. Negrello, and J. Wall. Radio and millimeter continuum surveys and their astrophysical implications. *Astron. Astrophys. Rev.*, 18:1–65, Feb. 2010. doi:10.1007/s00159-009-0026-0.

L. A. Díaz-García et al. Stellar populations of galaxies in the ALHAMBRA survey up to $z \sim 1$. *A&A*, 582:A14, Sep 2015. doi:10.1051/0004-6361/201425582.

J. M. Diego, P. Protopapas, H. B. Sandvik, and M. Tegmark. Non-parametric inver-

sion of strong lensing systems. *MNRAS*, 360:477–491, 2005a. doi:10.1111/j.1365-2966.2005.09021.x.

J. M. Diego, H. B. Sandvik, P. Protopapas, M. Tegmark, et al. Non-parametric mass reconstruction of A1689 from strong lensing data with the Strong Lensing Analysis Package. *MNRAS*, 362:1247–1258, 2005b. doi:10.1111/j.1365-2966.2005.09372.x.

J. M. Diego, M. Tegmark, P. Protopapas, and H. B. Sandvik. Combined reconstruction of weak and strong lensing data with WSLAP. *MNRAS*, 375:958–970, 2007. doi:10.1111/j.1365-2966.2007.11380.x.

J. M. Diego, T. Broadhurst, A. Zitrin, D. Lam, J. Lim, et al. Hubble Frontier Field free-form mass mapping of the massive multiple-merging cluster MACSJ0717.5+3745. *MNRAS*, 451(4):3920–3932, Aug 2015. doi:10.1093/mnras/stv1168.

J. M. Diego, T. Broadhurst, C. Chen, J. Lim, A. Zitrin, et al. A free-form prediction for the reappearance of supernova Refsdal in the Hubble Frontier Fields cluster MACSJ1149.5+2223. *MNRAS*, 456(1):356–365, Feb 2016. doi:10.1093/mnras/stv2638.

J. M. Diego, K. B. Schmidt, T. Broadhurst, D. Lam, J. Vega-Ferrero, et al. A free-form lensing model of A370 revealing stellar mass dominated BCGs, in Hubble Frontier Fields images. *MNRAS*, 473(4):4279–4296, Feb 2018a. doi:10.1093/mnras/stx2609.

J. M. Diego et al. Dark Matter under the Microscope: Constraining Compact Dark Matter with Caustic Crossing Events. *ApJ*, 857:25, Apr. 2018b. doi:10.3847/1538-4357/aab617.

S. Dieleman, K. W. Willett, and J. Dambre. Rotation–invariant convolutional neural networks for galaxy morphology prediction. *MNRAS*, 450(2):1441–1459, Apr 2015. doi:10.1093/mnras/stv632.

H. Domínguez Sánchez, M. Huertas-Company, M. Bernardi, D. Tuccillo, and J. L. Fischer. Improving galaxy morphologies for SDSS with Deep Learning. *MNRAS*, 476(3):3661, Feb 2018. doi:10.1093/mnras/sty338.

M. Donahue, G. Mark Voit, A. Mahdavi, K. Umetsu, S. Ettori, J. Merten, and M. Postman. CLASH-X: A Comparison of Lensing and X-ray Techniques for Measuring the Mass Profiles of Galaxy Clusters. *ApJ*, 794:136+157, 2014. doi:10.1088/0004-637X/794/2/136.

D. Donevski et al. Towards a census of high-redshift dusty galaxies with Herschel. A selection of "500 μm-risers". *A&A*, 614:A33, June 2018. doi:10.1051/0004-6361/201731888.

D. Dorfman and E. Alf. Maximum likelihood estimation of parameters of signal detection theory. *Psychometrika*, 33:117–123, 1968. doi:10.1007/BF02289677.

S. P. Driver, P. Norberg, I. K. Baldry, S. P. Bamford, A. M. Hopkins, J. Liske, J. Loveday, and J. A. Peacock. GAMA: towards a physical understanding of galaxy formation. *Astronom. Geophysics*, 50:12, 2009. doi:10.1111/j.1468-4004.2009.50512.x.

S. P. Driver, A. H. Wright, S. K. Andrews, L. J. Davies, P. R. Kafle, et al. Galaxy And Mass Assembly (GAMA): Panchromatic Data Release (far-UV – far-IR) and the low-z energy budget. *MNRAS*, 455:3911, 2016. doi:10.1093/mnras/stv2505.

S. Dye, C. Furlanetto, A. M. Swinbank, C. Vlahakis, J. W. Nightingale, et al. Revealing the complex nature of the strong gravitationally lensed system H-ATLAS J090311.6+003906 using ALMA. *MNRAS*, 452:2258, 2015. doi:10.1093/mnras/stv1442.

S. Eales et al. The Herschel ATLAS. *PASP*, 122:499, May 2010. doi:10.1086/653086.

S. A. Eales. Practical cosmology with lenses. *MNRAS*, 446:3224–3234, Jan. 2015. doi:10.1093/mnras/stu2214.

H. Ebeling, A. C. Edge, and J. P. Henry. MACS: A quest for the most massive galaxy clusters in the universe. *ApJ*, 553:668–676, 2001. doi:10.1086/320958.

H. Ebeling, E. Barrett, D. Donovan, C. J. Ma, A. C. Edge, and L. van Speybroeck. A Complete Sample of 12 Very X-Ray Luminous Galaxy Clusters at $z > 0.5$. *ApJ*, 661: L33, 2007. doi:10.1086/518603.

H. Ebeling, C. M, J. P. Kneib, E. Jullo, N. Courtney, E. Barrett, A. Edge, and J.-F. Le Borgne. A spectacular giant arc in the massive cluster lens MACS J1206.2–0847. *MNRAS*, 395:1213–1224, 2009. doi:10.1111/j.1365-2966.2009.14502.x.

H. Ebeling, A. C. Edge, A. Mantz, E. Barrett, J. Henry, C. J. Ma, and L. van Speybroeck. The X-ray brightest clusters of galaxies from the Massive Cluster Survey. *MNRAS*, 407: 83–93, 2010. doi:10.1111/j.1365-2966.2010.16920.x.

T. Eichner, S. Setiz, S. H. Suyu, A. Halkola, K. Umetsu, A. Zitrin, et al. Galaxy halo truncation and Giant Arc Surface Brightness Reconstruction in the Cluster MACS J1206.2–0847. *ApJ*, 774:124–158, 2013. doi:10.1088/0004-637X/774/2/124.

M. Eriksen et al. The PAU Survey. *MNRAS*, 484:4200–4215, Jan 2019. doi:10.1093/mnras/stz204.

J. Estrada et al. A Systematic Search for High Surface Brightness Giant Arcs in a Sloan Digital Sky Survey Cluster Sample. *ApJ*, 660(2):1176–1185, May 2007. doi:10.1086/512599.

E. E. Falco, C. S. Kochanek, and J. A. Muñoz. Limits on Cosmological Models from Radio-selected Gravitational Lenses. *ApJ*, 494(1):47–59, Feb. 1998. doi:10.1086/305207.

L. Fei-Fei, K. Andrej, and J. Justin. Cs231n convolutional neural networks for visual recognition. https://cs231n.github.io/, 2020. Accessed: 2021-02-18.

H. Fu et al. A comprehensive view of a strongly lensed Planck-associated submillimeter galaxy. *ApJ*, 753(2):134–12, June 2012. doi:10.1088/0004-637X/753/2/134.

Y. Fudamoto, R. J. Ivison, I. Oteo, M. Krips, Z. Y. Zhang, et al. The most distant, luminous, dusty star-forming galaxies: redshifts from NOEMA and ALMA spectral scans. *MNRAS*, 472:2028–2041, 2017. doi:10.1093/mnras/stx1956.

K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202, 1980. doi:10.1007/BF00344251.

R. Gavazzi, Y. Mellier, B. Fort, J. C. Cuillandre, and M. Dantel-Fort. Mass and light in the supercluster of galaxies MS0302+17. *A&A*, 422:407–422, Aug. 2004. doi:10.1051/0004-6361:20047109.

R. Gavazzi, P. J. Marshall, T. Treu, and A. Sonnenfeld. RingFinder: Automated Detection of Galaxy-scale Gravitational Lenses in Ground-based Multi-filter Imaging Data. *ApJ*, 785(2):144, Apr. 2014. doi:10.1088/0004-637X/785/2/144.

R. Gençay. Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging. *IEEE Transactions on Neural Networks*, 12: 726, 2001. doi:10.1109/72.935086.

M. Girardi et al. CLASH-VLT: Substructure in the galaxy cluster MACS J1206.2–0847 from kinematics of galaxy populations. *A&A*, 579:A4, 2015. doi:10.1051/0004-6361/201425599.

X. Glorot, A. Bordes, and Y. Bengio. Deep Sparse Rectifier Neural Networks. *Proceedings of Machine Learning Research*, 15:315–323, 2011. http://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf.

G. Golse and J. P. Kneib. Pseudo elliptical lensing mass model: Application to the NFW mass distribution. *A&A*, 390:821, 2002. doi:10.1051/0004-6361:20020639.

J. González-Nuevo, F. Argüeso, M. López-Caniego, L. Toffolatti, J. L. Sanz, P. Vielva, and D. Herranz. The Mexican hat wavelet family: application to point-source detection in cosmic microwave background maps. *MNRAS*, 369:1603–1610, July 2006. doi:10.1111/j.1365-2966.2006.10442.x.

J. González-Nuevo, A. Lapi, S. Fleuren, S. Bressan, L. Danese, G. De Zotti, and M. Ne-

grello. Herschel-ATLAS: Towards a sample of $\sim$ 1000 strongly lensed galaxies. *ApJ*, 749: 65–77, 2012. doi:10.1088/0004-637X/749/1/65.

J. González-Nuevo et al. Herschel-ATLAS/GAMA: SDSS cross-correlation induced by weak lensing. *MNRAS*, 442:2680–2690, Aug. 2014. doi:10.1093/mnras/stu1041.

J. González-Nuevo et al. H-ATLAS/GAMA: magnification bias tomography. Astrophysical constraints above $\sim$1 arcmin. *JCAP*, 10:024, Oct. 2017. doi:10.1088/1475-7516/2017/10/024.

I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout Networks. volume 28 (3), page 1319, 2013. arXiv:1302.4389v4.

J. Greenslade et al. Candidate high-z protoclusters among the Planck compact sources, as revealed by Herschel-SPIRE. *MNRAS*, 476:3336–3359, May 2018. doi:10.1093/mnras/sty023.

M. J. Griffin et al. The Herschel-SPIRE instrument and its in-flight performance. *A&A*, 518:L3, July 2010. doi:10.1051/0004-6361/201014519.

R. H. R. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405 (6789):947, 2000. doi:10.1038/35016072.

P. Hála. Spectral classification using convolutional neural networks. Master's thesis, 2014. arXiv:1412.8341v1.

A. Halkola, S. Seitz, and M. Pannella. Parametric strong gravitational lensing analysis of Abell 1689. *MNRAS*, 372(4):1425–1462, 10 2006. doi:10.1111/j.1365-2966.2006.10948.x.

P. Hartley, R. Flamary, N. Jackson, A. S. Tagore, and R. B. Metcalf. Support vector machine classification of strong gravitational lenses. *MNRAS*, 471(3):3378–3397, Nov. 2017. doi:10.1093/mnras/stx1733.

K. He, X. Zhang, S. Ren, and J. Sun. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *Proceedings of the 2015 IEEE International Conference on Computer Vision*, page 1026, 2015. doi:10.1109/ICCV.2015.123.

D. Herranz and J. L. Sanz. Matrix Filters for the Detection of Extragalactic Point Sources in Cosmic Microwave Background Images. *IEEE Journal of Selected Topics in Signal Processing*, 2:727–734, Nov. 2008. doi:10.1109/JSTSP.2008.2005339.

D. Herranz and P. Vielva. Cosmic microwave background images. *IEEE Signal Processing Magazine*, 27:67–75, Jan. 2010. doi:10.1109/MSP.2009.934716.

D. Herranz, J. L. Sanz, M. P. Hobson, R. B. Barreiro, J. M. Diego, E. Martínez-González, and A. N. Lasenby. Filtering techniques for the detection of Sunyaev-Zel′dovich clusters in multifrequency maps. *MNRAS*, 336:1057, 2002. doi:10.1046/j.1365-8711.2002.05704.x.

D. Herranz, M. López-Caniego, J. L. Sanz, and J. González-Nuevo. A novel multifrequency technique for the detection of point sources in cosmic microwave background maps. *MNRAS*, 394:510–520, Mar. 2009. doi:10.1111/j.1365-2966.2008.14336.x.

D. Herranz, F. Argüeso, and P. Carvalho. Compact Source Detection in Multichannel Microwave Surveys: From SZ Clusters to Polarized Sources. *Advances in Astronomy*, 2012:410965, Jan. 2012. doi:10.1155/2012/410965.

D. Herranz et al. Herschel-ATLAS: Planck sources in the phase 1 fields. *A&A*, 549:A31, Jan. 2013. doi:10.1051/0004-6361/201219435.

Y. D. Hezaveh, L. P. Levasseur, and P. J. Marshall. Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, 548(7669):555–557, Aug 2017. doi:10.1038/nature23463.

S. Hilbert, L. Marian, R. E. Smith, and V. Desjacques. Measuring primordial non-Gaussianity with weak lensing surveys. *MNRAS*, 426(4):2870–2888, Oct 2012. doi:10.1111/j.1365-2966.2012.21841.x.

M. Ho, M. M. Rau, M. Ntampaka, A. Farahi, H. Trac, and B. Póczos. A Robust and Efficient Deep Learning Method for Dynamical Mass Measurements of Galaxy Clusters. *ApJ*, 887(1):25, Dec 2019. doi:10.3847/1538-4357/ab4f82.

M. T. Hogan, B. R. McNamara, F. Pulido, P. E. J. Nulsen, H. R. Russell, et al. Mass Distribution in Galaxy Cluster Cores. *ApJ*, 837:51, 2017. doi:10.3847/1538-4357/aa5f56.

A. Horesh, E. O. Ofek, D. Maoz, M. Bartelmann, M. Meneghetti, and H.-W. Rix. The Lensed Arc Production Efficiency of Galaxy Clusters: A Comparison of Matched Observed and Simulated Samples. *ApJ*, 633(2):768–780, Nov. 2005. doi:10.1086/466519.

B. Hoyle. Measuring photometric redshifts using galaxy images and Deep Neural Networks. *Astronomy and Computing*, 16:34–40, July 2016. doi:10.1016/j.ascom.2016.03.006.

X. Huang, C. Storfer, V. Ravi, A. Pilon, et al. Finding Strong Gravitational Lenses in the DESI DECam Legacy Survey. *ApJ*, 894(1):78, May 2020. doi:10.3847/1538-4357/ab7ffb.

M. Huertas-Company et al. A Catalog of Visual-like Morphologies in the 5 Candels Fields using Deep Learning. *ApJS*, 221(1):8, Oct 2015. doi:10.1088/0067-0049/221/1/8.

D. Huterer and D. L. Shafer. Dark energy two decades after: observables, probes, consistency tests. *Rep. Prog. Phys.*, 81(1):016901, Dec 2017. doi:10.1088/1361-6633/aa997e.

S. Ikarashi, R. J. Ivison, K. I. Caputi, K. Nakanishi, C. D. P. Lagos, M. L. N. Ashby, et al. Extremely Red Submillimeter Galaxies: New $z \gtrsim$ 4-6 Candidates Discovered using ALMA and Jansky VLA. *ApJ*, 835:286, Feb. 2017. doi:0.3847/1538-4357/835/2/286.

N. Inada et al. The Sloan Digital Sky Survey Quasar Lens Search. V. Final Catalog from the Seventh Data Release. *AJ*, 143(5):119, May 2012. doi:10.1088/0004-6256/143/5/119.

M. J. Irwin, R. L. Webster, P. C. Hewett, R. T. Corrigan, and R. I. Jedrzejewski. Photometric Variations in the Q2237+0305 System: First Detection of a Microlensing Event. *AJ*, 98:1989, Dec. 1989. doi:10.1086/115272.

R. Ivison, A. M. Swinbank, I. Smail, A. I. Harris, R. S. Bussmann, A. Cooray, P. Cox, H. Fu, and A. Kovacs. Herschel-ATLAS: a binary HyLIRG pinpointing a cluster of starbursting proto-ellipticals. *ApJ*, 772:137, 2013. doi:10.1088/0004-637X/772/2/137.

R. J. Ivison, A. M. Swinbank, and B. Swinyard. Herschel and SCUBA-2 imaging and spectroscopy of a bright, lensed submillimetre galaxy at $z = 2.3$. *A&A*, 518:L35, 2010. doi:10.1051/0004-6361/201014548.

R. J. Ivison et al. The Space Density of Luminous Dusty Star-forming Galaxies at $z > 4$. *ApJ*, 832:78, Nov. 2016. doi:10.3847/0004-637X/832/1/78.

C. Jacobs, K. Glazebrook, T. Collett, A. More, and C. McCarthy. Finding strong lenses in CFHTLS using convolutional neural networks. *MNRAS*, 471(1):167–181, Jun 2017. doi:10.1093/mnras/stx1492.

C. Jacobs, T. Collett, K. Glazebrook, E. Buckley-Geer, H. T. Diehl, et al. An Extended Catalog of Galaxy-Galaxy Strong Gravitational Lenses Discovered in DES Using Convolutional Neural Networks. *ApJS*, 243(1):17, July 2019. doi:10.3847/1538-4365/ab26b6.

C. Jacobs et al. Finding high-redshift strong lenses in DES using convolutional neural networks. *MNRAS*, 484(4):5330–5349, Jan 2019. doi:10.1093/mnras/stz272.

Y. P. Jing and Y. Suto. The Density Profiles of the Dark Matter Halo Are Not Universal. *ApJ*, 529:L69–L72, 2000. doi:10.1086/312463.

R. Joseph et al. A PCA-based automated finder for galaxy-scale strong lenses. *A&A*, 566: A63, June 2014. doi:10.1051/0004-6361/201423365.

M. Kaplinghat, S. Tulin, and H.-B. Yu. Direct Detection Portals for Self-interacting Dark Matter. *Physical Review Letters*, 89:18, 2013. doi:10.1103/PhysRevD.89.035009.

M. Kaplinghat, R. E. Keeley, T. Linden, and H.-B. Yu. Tying Dark Matter to Baryons with Self-interactions. *Physical Review Letters*, 113:5, 2014. doi:10.1103/PhysRevLett.113.021302.

P. L. Kelly, J. M. Diego, S. Rodney, et al. Extreme magnification of an individual star at redshift 1.5 by a galaxy-cluster lens. *Nature Astronomy*, 2(4):334–342, 2018. doi:10.1038/s41550-018-0430-3.

E. J. Kim and R. J. Brunner. Star-galaxy classification using deep convolutional neural networks. *MNRAS*, 464(4):4463–4475, Feb. 2017. doi:10.1093/mnras/stw2672.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *a conference paper published at ICLR*, 2015. arXiv:1412.6980.

J. P. Kneib. In *The Shapes of Galaxies and Their Dark Halos*, page 50, New Haven, Connecticut, USA, 2002. Natarajan, P., ed. doi:10.1142/4886 .

J. P. Kneib, Y. Mellier, B. Fort, and G. Mathez. The distribution of dark matter in distant cluster-lenses: modelling A 370. *A&A*, 273:367, June 1993. 1993A&A...273..367K.

J. P. Kneib, Y. Mellier, R. Pello, J. Miralda-Escude, J. F. Le Borgne, H. Boehringer, and J. P. Picat. Dynamics of Abell 2218 from optical and near-IR imagery of arc(let)s and the ROSAT/HRI X-ray map. *A&A*, 303:27, Nov. 1995. arXiv:astro-ph/9504038.

J. P. Kneib, R. S. Ellis, I. Smail, W. J. Couch, and R. M. Sharples. Hubble Space Telescope Observations of the Lensing Cluster Abell 2218. *ApJ*, 471:643, Nov. 1996. doi:10.1086/177995.

C. S. Kochanek. The Analysis of Gravitational Lens Surveys. II. Maximum Likelihood Models and Singular Potentials. *ApJ*, 419:12, Dec. 1993. doi:10.1086/173455.

C. S. Kochanek. Is There a Cosmological Constant? *ApJ*, 466:638, Aug. 1996. doi:10.1086/177538.

C. S. Kochanek and P. L. Schechter. The Hubble Constant from Gravitational Lens Time Delays. In W. L. Freedman, editor, *Measuring and Modeling the Universe*, page 117, Jan. 2004. arXiv:astro-ph/0306040v1.

L. V. E. Koopmans, T. Treu, C. D. Fassnacht, R. D. Blandford, and G. Surpi. The Hubble Constant from the Gravitational Lens B1608+656. *ApJ*, 599(1):70–85, Dec. 2003. doi:10.1086/379226.

A. V. Kravtsov and S. Borgani. Formation of Galaxy Clusters. *Annu. Rev. Astron. Astrophys.*, 50:353–409, Sept. 2012. doi:10.1146/annurev-astro-081811-125502.

M. Kriek and C. Conroy. The Dust Attenuation Law in Distant Galaxies: Evidence for Variation with Spectral Type. 775:L16, 2013. doi:10.1088/2041-8205/775/1/L16.

M. Kriek, P. G. van Dokkum, I. Labbé, M. Franx, and G. D. Illingworth. An Ultra-Deep Near-Infrared Spectrum of a Compact Quiescent Galaxy at $z = 2.2$. *ApJ*, 700:221–231, 2009. doi:10.1088/0004-637X/700/1/221.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep ConvolutionalNeural Networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems*, 1:1097, 2012. doi:10.1145/3065386.

J. M. Kubo and I. P. Dell'Antonio. A method to search for strong galaxy-galaxy lenses in optical imaging surveys. *MNRAS*, 385(2):918–928, Apr. 2008. doi:10.1111/j.1365-2966.2008.12880.x.

T. Kundić, E. L. Turner, W. N. Colley, I. Gott, J. R., J. E. Rhoads, et al. A Robust Determination of the Time Delay in 0957+561A, B and a Measurement of the Global Value of Hubble's Constant. *ApJ*, 482(1):75–82, June 1997. doi:10.1086/304147.

F. Kurpiel, R. Minetto, and B. Nassu. Convolutional Neural Networks for License Plate Detection in Images. September 2017. doi:10.1109/ICIP.2017.8296912.

F. Lanusse et al. CMU DeepLens: deep learning for automatic image-based galaxy-galaxy strong lens finding. *MNRAS*, 473(3):3895–3906, Jan. 2018. doi:10.1093/mnras/stx1665.

L. F. Lanz, D. Herranz, J. L. Sanz, J. González-Nuevo, and M. López-Caniego. A multi-frequency method based on the matched multifilter for the detection of point sources in CMB maps. *MNRAS*, 403:2120–2130, 2010. doi:10.1111/j.1365-2966.2010.16260.x.

L. F. Lanz, D. Herranz, M. López-Caniego, J. González-Nuevo, G. de Zotti, et al. Extra-galactic point source detection in Wilkinson Microwave Anisotropy Probe 7-year data at 61 and 94 GHz. *MNRAS*, 428:3048–3057, 2013. doi:10.1093/mnras/sts252.

A. Lapi, J. González-Nuevo, L. Fan, A. Bressan, G. De Zotti, L. Danese, et al. Herschel-ATLAS Galaxy Counts and High-redshift Luminosity Functions: The Formation of Massive Early-type Galaxies. *ApJ*, 742:24, Nov. 2011. doi:10.1088/0004-637X/742/1/24.

R. Laureijs et al. Euclid Definition Study Report. Oct. 2011. arXiv:1110.3193v1.

S. M. Leach et al. Component separation methods for the PLANCK mission. *A&A*, 491: 597–615, Nov. 2008. doi:10.1051/0004-6361:200810116.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning ap-

plied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi:10.1109/5.726791.

Y. LeCun, Y. Bengio, and G. Hinton. Deep Learning. *Nature*, 521:436–44, 05 2015. doi:10.1038/nature14539.

F. Lenzen, S. Schindler, and O. Scherzer. Automatic detection of arcs and arclets formed by gravitational lensing. *A&A*, 416:391–401, Mar. 2004. doi:10.1051/0004-6361:20034619.

A. Lewis and A. Challinor. Weak gravitational lensing of the CMB. *Phys. Rep.*, 429(1): 1–65, June 2006. doi:10.1016/j.physrep.2006.03.002.

J. Liesenborgs, S. De Rijcke, H. Dejonghe, and P. Bekaert. Non-parametric inversion of gravitational lensing systems with few images using a multi-objective genetic algorithm. *MNRAS*, 380(4):1729–1736, 09 2007. doi:10.1111/j.1365-2966.2007.12236.x.

M. Limousin, H. Ebeling, C. J. Ma, A. M. Swinbank, G. P. Smith, et al. MACS J1423.8+2404: gravitational lensing by a massive, relaxed cluster of galaxies at $z = 0.54$. *MNRAS*, 405(2):777–782, June 2010. doi:10.1111/j.1365-2966.2010.16518.x.

R. Link and M. J. Pierce. Cosmological Parameters from Multiple-Arc Gravitational Lensing Systems. I. *ApJ*, 502(1):63–74, July 1998. doi:10.1086/305892.

M. López-Caniego, D. Herranz, J. González-Nuevo, J. L. Sanz, R. B. Barreiro, P. Vielva, F. Argüeso, and L. Toffolatti. Comparison of filters for the detection of point sources in Planck simulations. *MNRAS*, 370:2047, 2006. doi:10.1111/j.1365-2966.2006.10639.x.

LSST Science Collaborations. LSST Science Book, Version 2.0, 2009. arXiv:0912.0201.

B. W. Lyke et al. The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release. *ApJS*, 250(1):8, Aug 2020. doi:10.3847/1538-4365/aba623.

S. J. Maddox, E. Valiante, P. Cigan, L. Dunne, S. Eales, et al. The Herschel-ATLAS Data Release 2. Paper II. *ApJS*, 236:30, June 2018. doi:10.3847/1538-4365/aab8fc.

A. Manjón-García, D. Herranz, J. M. Diego, L. Bonavera, and J. González-Nuevo. Multi-frequency filter search for high redshift sources and lensing systems in Herschel-ATLAS. *A&A*, 622:A106, Feb. 2019. doi:10.1051/0004-6361/201834549.

A. Manjón-García, J. M. Diego, D. Herranz, and D. Lam. Constraining the abundance of dark matter in the central region of the galaxy cluster MACS J1206.2–0847 with a free-form strong lensing analysis. *A&A*, 639:A125, July 2020. doi:10.1051/0004-6361/201936914.

D. Maoz and H. W. Rix. Early-Type Galaxies, Dark Halos, and Gravitational Lensing Statistics. *ApJ*, 416:425, Oct. 1993. doi:10.1086/173248.

M. Markevitch et al. Direct Constraints on the Dark Matter Self-Interaction Cross Section from the Merging Galaxy Cluster 1E 0657-56. *ApJ*, 606(2):819–824, May 2004. doi:10.1086/383178.

M. Maturi, S. Mizera, and G. Seidel. Multi-colour detection of gravitational arcs. *A&A*, 567:A111, July 2014. doi:10.1051/0004-6361/201321634.

J. McKean et al. Strong Gravitational Lensing with the SKA. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 84, Apr. 2015. arXiv:1502.03362.

M. Meneghetti. Introduction to Gravitational Lensing, 2016. https://www.ita.uni-heidelberg.de/~jmerten/misc/meneghetti_lensing.pdf.

M. Meneghetti, M. Bartelmann, and L. Moscardini. Cluster cross sections for strong lensing: analytic and numerical lens models. *MNRAS*, 346:67, 2003. doi:10.1046/j.1365-8711.2003.06276.x.

M. Meneghetti et al. The Frontier Fields lens modelling comparison project. *MNRAS*, 472: 3177–3216, 2017. doi:10.1093/mnras/stx2064.

J. Merten, M. Cacciato, M. Meneghetti, C. Mignone, and M. Bartelmann. Combining weak and strong cluster lensing: applications to simulations and MS 2137. *A&A*, 500: 681, 2009. doi:10.1051/0004-6361/200810372.

J. Merten, M. Meneghetti, M. Postman, K. Umetsu, et al. CLASH: The Concentration-Mass Relation of Galaxy Clusters. *ApJ*, 806:4, 2015. doi:10.1088/0004-637X/806/1/4.

H. Messias et al. Herschel-ATLAS and ALMA: HATLAS J142935.3-002836, a lensed major merger at redshift 1.027. *A&A*, 568:92, 2014. doi:10.1051/0004-6361/201424410.

R. B. Metcalf and M. Petkova. GLAMER Part I: A Code for Gravitational Lensing Simulations with Adaptive Mesh Refinement. *MNRAS*, 445(2):1942–1953, Oct 2014. doi:10.1093/mnras/stu1859.

R. B. Metcalf, M. Meneghetti, C. Avestruz, F. Bellagamba, et al. The strong gravitational lens finding challenge. *A&A*, 625:A119, May 2019. doi:10.1051/0004-6361/201832797.

M. J. Michałowski, C. C. Hayward, J. S. Dunlop, V. A. Bruce, M. Cirasuolo, et al. Determining the stellar masses of submillimetre galaxies: the critical importance of star formation histories. *A&A*, 571:A75, Nov. 2014. doi:10.1051/0004-6361/201424174.

MissingLink. https://missinglink.ai/guides/convolutional-neural-networks/convolutional-neural-network-tutorial-basic-advanced/. Accessed: 2021-02-18.

A. More, R. Cabanac, S. More, C. Alard, M. Limousin, et al. The CFHTLS-Strong Lensing Legacy Survey (SL2S): Investigating the Group-scale Lenses with the SARCS Sample. *ApJ*, 749(1):38, Apr. 2012. doi:10.1088/0004-637X/749/1/38.

W. R. Morningstar et al. Data-driven Reconstruction of Gravitationally Lensed Galaxies Using Recurrent Inference Machines. *ApJ*, 883(1):14, Sep 2019. doi:10.3847/1538-4357/ab35d7.

R. Narayan and M. Bartelmann. Lectures on Gravitational Lensing. *arXiv e-prints*, art. astro-ph/9606001, June 1996. arXiv:astro-ph/9606001v2.

P. Natarajan and J.-P. Kneib. Lensing by galaxy haloes in clusters of galaxies. *MNRAS*, 287(4):833–847, June 1997. doi:10.1093/mnras/287.4.833.

J. F. Navarro, C. S. Frenk, and S. D. M. White. The Structure of Cold Dark Matter Halos. *ApJ*, 462:563, 1996. doi:10.1086/177173.

J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. *ApJ*, 490:493, 1997. doi:10.1086/304888.

H. Nayyeri et al. Candidate Gravitationally Lensed Dusty Star-forming Galaxies in the Herschel Wide Area Surveys. *ApJ*, 823:17, 2016. doi:10.3847/0004-637X/823/1/17.

M. Negrello et al. Astrophysical and cosmological information from large-scale submm surveys of extragalactic sources. *MNRAS*, 377:1557, 2007. doi:10.1111/j.1365-2966.2007.11708.x.

M. Negrello et al. The Detection of a Population of Submillimeter-Bright, Strongly Lensed Galaxies. *Science*, 330:800, 2010. doi:10.1126/science.1193420.

M. Negrello et al. Herschel-ATLAS: deep HST/WFC3 imaging of strongly lensed submillimeter galaxies. *MNRAS*, 440:1999, 2014. doi:10.1093/mnras/stu413.

M. Negrello et al. The Herschel-ATLAS: a sample of 500 μm-selected lensed galaxies over 600 deg^2. *MNRAS*, 465:3558–3580, 2017. doi:10.1093/mnras/stw2911.

A. B. Newman et al. The Distribution of Dark Matter Over 3 Decades in Radius in the Lensing Cluster Abell 611. *ApJ*, 706:1078, 2009. doi:10.1088/0004-637X/706/2/1078.

OGLE Collaboration. OGLE 2003-BLG-235/MOA 2003-BLG-53: A Planetary Microlensing Event. *ApJ*, 606(2):L155–L158, May 2004. doi:10.1086/420928.

M. Oguri and P. J. Marshall. Gravitationally lensed quasars and supernovae in future wide-field optical imaging surveys. *MNRAS*, 405(4):2579–2593, July 2010. doi:10.1111/j.1365-2966.2010.16639.x.

I. Oteo et al. Witnessing the Birth of the Red Sequence. *ApJ*, 827:34, Aug. 2016. doi:10.3847/0004-637X/827/1/34.

B. Paczynski. Gravitational Microlensing by the Galactic Halo. *ApJ*, 304:1, May 1986. doi:10.1086/164140.

N. Palanque-Delabrouille et al. The extended Baryon Oscillation Spectroscopic Survey: Variability selection and quasar luminosity function. *A&A*, 587:A41, 2016. doi:10.1051/0004-6361/201527392.

C. J. Parde et al. Face and Image Representation in Deep CNN Features. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pages 673–680, 2017. doi:10.1109/FG.2017.85.

I. Pâris et al. The Sloan Digital Sky Survey Quasar Catalog: Twelfth data release. *A&A*, 597:A79, 2017. doi:10.1051/0004-6361/201527999.

I. Pâris et al. The Sloan Digital Sky Survey Quasar Catalog: Fourteenth Data Release. *A&A*, 613:A51, 2018. doi:.

J. Pasquet, E. Bertin, M. Treyer, S. Arnouts, and D. Fouchez. Photometric redshifts from SDSS images using a convolutional neural network. *A&A*, 621:A26, Dec 2018. doi:10.1051/0004-6361/201833617.

R. S. Pawase, F. Courbin, C. Faure, R. Kokotanekova, and G. Meylan. A 7 deg^2 survey for galaxy-scale gravitational lenses with the HST imaging archive. *MNRAS*, 439(4): 3392–3404, Feb 2014. doi:10.1093/mnras/stu179.

E. A. Pearson et al. H-ATLAS: estimating redshifts of Herschel sources from sub-mm fluxes. *MNRAS*, 435:2753–2763, 2013. doi:10.1093/mnras/stt1369.

S. Perlmutter et al. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *ApJ*, 517(2):565–586, June 1999. doi:10.1086/307221.

L. Perreault Levasseur, Y. D. Hezaveh, and R. H. Wechsler. Uncertainties in Parameters Estimated with Neural Networks: Application to Strong Gravitational Lensing. *ApJ*, 850(1):L7, Nov 2017. doi:10.3847/2041-8213/aa9704.

C. E. Petrillo et al. Finding strong gravitational lenses in the Kilo Degree Sur-

vey with Convolutional Neural Networks. *MNRAS*, 472(1):1129–1150, Aug 2017. doi:10.1093/mnras/stx2052.

C. E. Petrillo et al. LinKS: Discovering galaxy-scale strong lenses in the Kilo-Degree Survey using convolutional neural networks. *MNRAS*, 484(3):3879–3896, Jan 2019. doi:10.1093/mnras/stz189.

C. E. o. Petrillo. Testing convolutional neural networks for finding strong gravitational lenses in KiDS. *MNRAS*, 482(1):807–820, Jan. 2019. doi:10.1093/mnras/sty2683.

M. M. Pieri et al. Weave-qso: A massive intergalactic medium survey for the william herschel telescope, 2016. arXiv:1611.09388.

G. L. Pilbratt et al. Herschel Space Observatory. An ESA facility for far-infrared and submillimetre astronomy. *A&A*, 518:L1, July 2010. doi:10.1051/0004-6361/201014759.

Planck Collaboration. Planck Early Results. VII. The Early Release Compact Source Catalogue. *A&A*, 536:A7, Dec. 2011. doi:10.1051/0004-6361/201116474.

Planck Collaboration. Planck 2013 results. XXVIII. The Planck Catalogue of Compact Sources. *A&A*, 571:A28, Nov. 2014. doi:10.1051/0004-6361/201321524.

Planck Collaboration. Planck 2015 results. XXVI. The Second Planck Catalogue of Compact Sources. *A&A*, 594:A26, Sept. 2016. doi:10.1051/0004-6361/201526914.

Planck Collaboration. Planck intermediate results. LV. The Planck Multi-frequency Catalogue of Non-thermal Sources. *A&A*, Feb. 2018. doi:10.1051/0004-6361/201832888.

Planck Collaboration. Planck 2018 results. *A&A*, 641:A6, Sep 2020. doi:10.1051/0004-6361/201833910.

M. Postman, D. Coe, N. Benítez, L. Bradley, and T. Broadhurst. Cluster Lensing And Supernova survey with Hubble (CLASH): An Overview. *ApJS*, 199:25, 2012. doi:10.1088/0067-0049/199/2/25.

M. Pourrahmani, H. Nayyeri, and A. Cooray. LensFlow: A Convolutional Neural Network in Search of Strong Gravitational Lenses. *ApJ*, 856(1):68, Mar 2018. doi:10.3847/1538-4357/aaae6a.

D. M. W. Powers. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *J. Mach. Learn. Technol*, 2:2229–3981, 01 2011. doi:10.9735/2229-3981.

L. Pozzetti, M. Bolzonella, E. Zucca, G. Zamorani, S. Lilly, et al. zCOSMOS - 10k-bright spectroscopic sample. *A&A*, 523:A13, Nov. 2010. doi:10.1051/0004-6361/200913020.

L. Prechelt. *Neural Networks: Tricks of the Trade*. Springer, Berlin, Heidelberg, 1998. doi:10.1007/978-3-642-35289-8.

S. Rawlings and R. Schilizzi. The Square Kilometre Array, 2011. arXiv:1105.5953v1.

A. Refregier, A. Amara, T. D. Kitching, A. Rassat, R. Scaramella, and J. Weller. Euclid Imaging Consortium Science Book, 2010. arXiv:1001.0061.

S. Refsdal. On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect. *MNRAS*, 128:307, Jan. 1964. doi:10.1093/mnras/128.4.307.

A. G. Riess et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116(3):1009–1038, Sep 1998. doi:10.1086/300499.

R. Robinson. Convolutional neural networks - basics. `https://mlnotebook.github.io/post/CNN1/`, 2017. Accessed: 2021-02-18.

M. Rocha et al. Cosmological Simulations with Self-Interacting Dark Matter I: Constant Density Cores and Substructure. *MNRAS*, 430:81–104, 2012. doi:10.1093/mnras/sts514.

P. Rosati, I. Balestra, C. Grillo, A. Mercurio, M. Nonino, et al. CLASH-VLT: A VIMOS Large Programme to Map the Dark Matter Mass Distribution in Galaxy Clusters and Probe Distant Lensed Galaxies. *The Messenger*, 158:48, 2014. 2014Msngr.158...48R.

O. Russakovsky et al. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)*, 115:211, 2015. doi:10.1007/s11263-015-0816-y.

E. S. Rykoff, E. Rozo, M. T. Busha, C. E. Cunha, et al. redMaPPer I: Algorithm and SDSS DR8 Catalog. *ApJ*, 785:104–137, 2014. doi:10.1088/0004-637X/785/2/104.

P. Saha and L. L. R. Williams. Non-parametric reconstruction of the galaxy lens in PG 1115+080. *MNRAS*, 292(1):148–156, 11 1997. doi:10.1093/mnras/292.1.148.

D. J. Sand, T. Treu, and R. S. Ellis. The Dark Matter Density Profile of the Lensing Cluster MS 2137-23. *ApJ*, 574(2):L129–L133, Aug. 2002. doi:10.1086/342530.

D. J. Sand, T. Treu, G. P. Smith, and R. Ellis. The Dark Matter Distribution in the Central Regions of Galaxy Clusters: Implications for Cold Dark Matter. *ApJ*, 604:88, 2004. doi:10.1086/382146.

C. Schaefer, M. Geiger, T. Kuntzer, and J. P. Kneib. Deep convolutional neural networks as strong gravitational lens detectors. *A&A*, 611:A2, Mar. 2018. doi:10.1051/0004-6361/201731201.

P. L. Schechter et al. The Quadruple Gravitational Lens PG 1115+080: Time Delays and Models. *ApJ*, 475(2):L85–L88, Feb. 1997. doi:10.1086/310478.

M. Schmidt. 3C 273 : A Star-Like Object with Large Red-Shift. *Nature*, 197(4872):1040, Mar. 1963. doi:10.1038/1971040a0.

P. Schneider. *Extragalactic Astronomy and Cosmology*. 2006. doi:10.1007/978-3540331759.

P. Schneider, J. Ehlers, and E. E. Falco. *Gravitational Lenses*. 1992. doi:10.1007/978-3662037584.

SDSS-III collaboration. The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III. *ApJ*, 193:29–50, 2011. doi:10.1088/0067-0049/193/2/29.

G. Seidel and M. Bartelmann. Arcfinder: an algorithm for the automatic detection of gravitational arcs. *A&A*, 472(1):341–352, Sept. 2007. doi:10.1051/0004-6361:20066097.

I. Sendra, J. M. Diego, T. Broadhurst, and R. Lazkoz. Enabling non-parametric strong lensing models to derive reliable clustermass distributions - WSLAP+. *MNRAS*, 437: 2642–2651, 2014. doi:10.1093/mnras/stt2076.

A. L. Serra and M. J. L. Domínguez Romero. Measuring the dark matter equation of state. *MNRAS*, 415(1):L74–L77, July 2011. doi:10.1111/j.1745-3933.2011.01082.x.

F. Sha, L. Saul, and D. Lee. Multiplicative updates for nonnegative quadratic programming in support vector machines. In *Advances in Neural Information Processing Systems 15*, pages 1065–1072. MIT Press, 2003. doi:10.1162/neco.2007.19.8.2004.

Y. Shu et al. The Sloan Lens ACS Survey. XIII. Discovery of 40 New Galaxy-scale Strong Lenses. *ApJ*, 851(1):48, Dec 2017. doi:10.3847/1538-4357/aa9794.

M. W. L. Smith, E. Ibar, S. J. Maddox, et al. The Herschel-ATLAS Data Release 2, Paper I. *ApJ*, 233:21, 2017. doi:10.3847/1538-4365/aa9b35.

J. Spilker, D. Marrone, M. Aravena, M. Bethermin, M. Bothwell, et al. ALMA Imaging and Gravitational Lens Models of South Pole Telescope-Selected Dusty, Star-Forming Galaxies at High Redshifts. *ApJ*, 826:112, 2016. doi:10.3847/0004-637X/826/2/112.

N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929, 2014. doi:10.5555/2627435.2670313.

D. P. Stark et al. The CASSOWARY spectroscopy survey: a new sample of gravitationally lensed galaxies in SDSS. *MNRAS*, 436(2):1040–1056, Dec. 2013. doi:10.1093/mnras/stt1624.

M. L. Strandet et al. ISM Properties of a Massive Dusty Star-forming Galaxy Discovered at $z \sim 7$. *ApJ*, 842:L15, June 2017. doi:10.3847/2041-8213/aa74b0.

J. A. Tauber, N. Mandolesi, J. L. Puget, et al. Planck pre-launch status: The Planck mission. *A&A*, 520:A1, Sept. 2010. doi:10.1051/0004-6361/200912983.

M. Tegmark and A. de Oliveira-Costa. Removing Point Sources From CMB Maps. *ApJ*, 500:L83, 1998. doi:10.1086/311410.

The Dark Energy Survey Collaboration. The Dark Energy Survey. *arXiv e-prints*, Oct. 2005. arXiv:astro-ph/0510346.

T. Treu, A. A. Dutton, M. W. Auger, P. J. Marshall, A. S. Bolton, et al. The SWELLS survey – I. A large spectroscopically selected sample of edge-on late–type lens galaxies. *MNRAS*, 417(3):1601–1620, Sep 2011. doi:10.1111/j.1365-2966.2011.19378.x.

C. S. Trotter, J. N. Winn, and J. N. Hewitt. A Multipole-Taylor Expansion for the Potential of the Gravitational Lens MG J0414+0534. *ApJ*, 535(2):671–691, June 2000. doi:10.1086/308861.

Y. Tsapras. Microlensing Searches for Exoplanets. *Geosciences*, 8(10):365, Sep 2018. doi:10.3390/geosciences8100365.

E. L. Turner, J. P. Ostriker, and I. Gott, J. R. The statistics of gravitational lenses: the distributions of image angular separations and lens redshifts. *ApJ*, 284:1–22, Sept. 1984. doi:10.1086/162379.

K. Umetsu. Cluster Weak Gravitational Lensing. Proceedings of the International School of Physics "Enrico Fermi", pages 269–300, Feb. 2011. doi:10.3254/978-1-60750-819-9-269.

K. Umetsu, T. Broadhurst, A. Zitrin, E. Medezinski, and L. Y. Shu. Cluster Mass Profiles from a Bayesian Analysis of Weak Lensing Distortion and Magnification Measurements: Applications to Subaru Data. *ApJ*, 729:127, 2011. doi:10.1088/0004-637X/729/2/127.

K. Umetsu et al. CLASH: Mass Distribution in and around MACS J1206.2-0847 from a Full Cluster Lensing Analysis. *ApJ*, 755:56, 2012. doi:10.1088/0004-637X/755/1/56.

E. Valiante et al. The Herschel-ATLAS data release 1 - I. Maps, catalogues and number counts. *MNRAS*, 462:3146–3179, 2016. doi:10.1093/mnras/stw1806.

E. van Uitert, M. Cacciato, H. Hoekstra, and R. Herbonnet. Evolution of the luminosity-to-halo mass relation of LRGs from a combined analysis of SDSS-DR10+RCS2. *A&A*, 579:A26, 2015. doi:10.1051/0004-6361/201525834.

C. Vanderriest, J. Schneider, G. Herpe, M. Chevreton, M. Moles, and G. Wlerick. The value of the time delay delta T (A,B) for the 'double' quasar 0957+561 from optical photometric monitoring. *A&A*, 215:1–13, May 1989. 1989A&A...215....1V.

J. Vega-Ferrero, J. M. Diego, V. Miranda, and G. M. Bernstein. The Hubble Constant from SN Refsdal. *ApJ*, 853(2):L31, Feb 2018. doi:10.3847/2041-8213/aaa95f.

A. Verma, T. Collett, G. P. Smith, et al. Strong Lensing considerations for the LSST observing strategy. *arXiv e-prints*, Feb. 2019. arXiv:1902.05141v1.

J. D. Vieira, D. P. Marrone, et al. Dusty starburst galaxies in the early Universe as revealed by gravitational lensing. *Nature*, 495:344, 2013. doi:10.1038/nature12001.

P. Vielva, E. Martínez-González, J. E. Gallegos, L. Toffolatti, and J. L. Sanz. Point source detection using the Spherical Mexican Hat Wavelet on simulated all-sky Planck maps. *MNRAS*, 344:89–104, Sept. 2003. doi:10.1046/j.1365-8711.2003.06792.x.

D. Walsh, R. F. Carswell, and R. J. Weymann. 0957+561 A, B: twin quasistellar objects or gravitational lens? *Nature*, 279:381–384, May 1979. doi:10.1038/279381a0.

J. Wambsganss. Gravitational Lensing in Astronomy. *Living Rev. in Relativity*, 1:12, 1998. doi:10.12942/lrr-1998-12.

J. Wambsganss. *Gravitational Lensing: Strong, Weak and Micro*. Springer Berlin Heidelberg, 2006. doi:10.1007/978-3-540-30310-7.

F. Wang et al. A luminous quasar at redshift 7.642. *ApJ*, 907(1):L1, Jan 2021. doi:110.3847/2041-8213/abd8c6.

L. Wang et al. HerMES: detection of cosmic magnification of submillimetre galaxies using angular cross-correlation. *MNRAS*, 414:596, 2011. doi:10.1111/j.1365-2966.2011.18417.x.

J. Wardlow et al. HerMES: Candidate Gravitationally Lensed Galaxies and Lensing Statistics at Submillimeter Wavelengths. *ApJ*, 762:59, 2013. doi:10.1088/0004-637X/762/1/59.

S. J. Warren and S. Dye. Semilinear Gravitational Lens Inversion. *ApJ*, 590(2):673–682, June 2003. doi:10.1086/375132.

S. Weinberg. Anthropic Bound on the Cosmological Constant. *Phys. Rev. Lett.*, 59:2607–2610, Nov 1987. doi:10.1103/PhysRevLett.59.2607.

K. W. Willett, C. J. Lintott, et al. Galaxy Zoo 2. *MNRAS*, 435(4):2835–2860, Nov. 2013. doi:10.1093/mnras/stt1458.

G. Wilson, S. Cole, and C. S. Frenk. Cluster mass reconstruction from weak gravitational lensing. *MNRAS*, 280(1):199–218, May 1996. doi:10.1093/mnras/280.1.199.

K. C. Wong et al. H0LiCOW - XIII. A 2.4 per cent measurement of H_0 from lensed quasars. *MNRAS*, 498:1420–1439, Jan. 2020. doi:10.1093/mnras/stz3094.

A. Wootten and A. Thompson. The Atacama Large Millimeter/Submillimeter Array. *Proceedings of the IEEE*, 97(8):1463–1471, Aug 2009. doi:10.1109/jproc.2009.2020572.

C. O. Wright and T. G. Brainerd. Gravitational Lensing by NFW Halos. *ApJ*, 534:54, 2000. doi:10.1086/308744.

J. S. B. Wyithe, E. L. Turner, and D. N. Spergel. Gravitational Lens Statistics for Generalized NFW Profiles. *ApJ*, 555:504–562, 2001. doi:10.1086/321437.

J. A. Zavala et al. A dusty star-forming galaxy at z=6 revealed by strong gravitational lensing. *Nature Astronomy*, 2:56, 2018. doi:10.1038/s41550-017-0297-8.

J. Zhang. Measuring the reduced shear. *JCAP*, 2011:41, Nov 2011. doi:10.1088/1475-7516/2011/11/041.

H. Zhao. Analytical models for galactic nuclei. *MNRAS*, 278:488–496, 1996. doi:10.1093/mnras/278.2.488.

A. Zitrin et al. CLASH: New Multiple-Images Constraining the Inner Mass Profile of MACS J1206.2-0847. *ApJ*, 749:97–104, 2012. doi:10.1088/0004-637X/749/2/97.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*, 67:301, 2005. doi:10.1111/j.1467-9868.2005.00503.x.

F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6: 110–127, Jan. 1933. doi:10.1007/s10714-008-0707-4.

F. Zwicky. Nebulae as Gravitational Lenses. *Phys. Rev.*, 51:290–290, Feb 1937. doi:10.1103/PhysRev.51.290.