scientific reports

Check for updates

OPEN

Leveraging a deep learning generative model to enhance recognition of minor asphalt defects

Saúl Cano-Ortiz[1,3]✉, Eugenio Sainz-Ortiz[1,3], Lara Lloret Iglesias[2,3], Pablo Martínez Ruiz del Árbol[2,3] & Daniel Castro-Fresno[1,3]

Deep learning-based computer vision systems have become powerful tools for automated and cost-effective pavement distress detection, essential for efficient road maintenance. Current methods focus primarily on developing supervised learning architectures, which are limited by the scarcity of annotated image datasets. The use of data augmentation with synthetic images created by generative models to improve these supervised systems is not widely explored. The few studies that do focus on generative architectures are mostly non-conditional, requiring extra labeling, and typically address only road crack defects while aiming to improve classification models rather than object detection. This study introduces AsphaltGAN, a novel class-conditional Generative Adversarial Network with attention mechanisms, designed to augment datasets with various rare road defects to enhance object detection. An in-depth analysis evaluates the impact of different loss functions and hyperparameter tuning. The optimized AsphaltGAN outperforms state-of-the-art generative architectures on public datasets. Additionally, a new workflow is proposed to improve object detection models using synthetic road images. The augmented datasets significantly improve the object detection metrics of You Only Look Once version 8 by 33.0%, 3.8%, 46.3%, and 51.8% on the Road Damage Detection 2022 dataset, Crack Dataset, Asphalt Pavement Detection Dataset, and Crack Surface Dataset, respectively.

Keywords Conditional generative model, Minor asphalt defect recognition, Data augmentation, Object detection, Road maintenance

As essential elements of civil infrastructure, asphalt pavements undergo degradation over time due to heavy traffic, unpredictable weather, poor construction practices, sub-optimal material quality, and inadequate maintenance[1]. The degradation of road conditions leads to the emergence of surface defects like cracks or potholes that worsen over time. Then, regular pavement inspection and proactive maintenance are crucial for ensuring driving safety and serviceability[2]. In fact, effective identification of asphalt pavement defects and timely maintenance measures at an early stage can significantly reduce economic expenditure and carbon emissions[3].

In recent years, the advancement of deep learning algorithms applied to computer vision tasks has attracted great attention in civil engineering, specifically in the field of automatic asphalt pavement distress recognition[4]. The advent of modern, cost-effective vehicle-mounted cameras and advancements in technological infrastructure have spurred research into computer-aided visual inspection techniques. The investigation focuses on three computer vision applications mainly using deep convolutional neural networks-based models: classification, segmentation, and object detection. Classification assigns a single label to the entire image, segmentation provides pixel-level classification, and object detection combines location of multiple defects (bounding boxes) and their categories (distress types).

Computer vision-based studies primarily concentrate on identifying common or most encountered types of asphalt road surface cracks (e.g., longitudinal, transverse, block, or alligator). However, the performance of deep learning-based algorithms is constrained by the limited ability to detect minor defects, which are also significant for road maintenance[5]. Concerning asphalt road damage detection, the more serious the distress, the

[1]GITECO Research Group, University of Cantabria, 39005 Santander, Spain. [2]Institute of Physics of Cantabria, UC-CSIC, 39005 Santander, Spain. [3]These authors contributed equally: Saúl Cano-Ortiz, Eugenio Sainz-Ortiz, Lara Lloret Iglesias, Pablo Martínez Ruiz del Árbol, and Daniel Castro-Fresno. ✉email: saul.cano@alumnos.unican.es; saulcano.ml@gmail.com

more difficult to collect the data. For instance, potholes, a prominent and severe road distress, are challenging to collect as they tend to be promptly repaired[6]. Also, sealed cracks are not usually analysed, and their recognition is crucial because they prevent water infiltration, which, when left unattended, can lead to structural damage and and a reduction in road service life[7]. There are also less common cracks, such as isolated cracks with diagonal/oblique and irregular geometry, often overlooked in research but are essential to address before their severity escalates. The initial challenge lies in identifying these minor defects to improve the robustness of detection models.

In real-world civil engineering scenarios, obtaining high-quality annotated large-scale data for long-tail defects is often challenging[8]. Deep learning-based classification, segmentation, and object detection algorithms are inherently data-driven, requiring balanced and sufficient data for optimal performance. Most studies propose complex architectures for pavement crack recognition, such as OUR-Net[9] or RHACrackNet[10], but they do not consider the importance of having a large volume of diverse, high-quality data. These limitations pose challenges in engineering practice, particularly in reducing overfitting and enhancing recognition[11]. A potential solution to these challenges is data augmentation, which can amplify the representation of rare asphalt pavement distresses. Traditional methods, such as affine transformations[12] (e.g., rotation or translation), are effective for datasets with balanced distributions but may only add minor artifacts to existing images without introducing new perspectives or content, thereby limiting intra-class diversity for tail images[13]. In contrast, generative models offer the capability to generate entirely new data that closely resembles the probability distribution of real images, surpassing traditional methods[14]. However, there is limited research focusing on applying generative models for synthetic data augmentation to improve supervised models, particularly in the context of object detection. In the following, an analysis of the current state-of-the-art will be conducted.

The most widely used generative architectures in the field of asphalt defect synthesis are Variational Autoencoders (VAE)[15] and Generative Adversarial Networks (GAN)[16]. The VAE uses an encoder network to map data to a latent space and a decoder to generate data from the latent space. A GAN consists of a generator network that produces fake data and a discriminator that distinguishes between real and generated data. Multiple enhanced GAN variants have been investigated such as Deep Convolutional GAN (DCGAN)[17], Wasserstein GAN (WGAN)[18], WGAN with Gradient Penalty (WGAN-GP)[19], Progressive GAN (PG-GAN)[20], super-resolution GAN (SRGAN)[21], conditional GAN (cGAN)[22], and cycle-consistent GAN (CycleGAN)[23].

Mazzini et al.[24] addressed the challenge of obtaining costly semantic segmentation ground truths by proposing a WGAN-GP-based method for augmenting datasets in semantic segmentation. Also, an iterative image retrieval system was proposed to match real images. Similarly, a method called ConnCrack[25] was proposed by exploring a WGAN. The generator of WGAN was designed to produce crack connectivity maps, while the discriminator checked if the maps and the original patch, were a real or a fake pair. Zhang et al.[26] proposed CrackGAN, a DCGAN trained with crack ground truth patches. The augmented dataset proved to improve the segmentation performance of an asymmetrical U-Net. Gao et al.[27] implemented a DCGAN with Leaf-Bootstrapping to generate synthetic frames to enhance a classification system. The LB method improves GAN performance by minimizing intra-class variation, allowing the generation of higher-quality images and ensuring better training stability. Pei et al.[28] trained a VAE encoder and utilized its pre-trained encoder to generate a feature map, which was then connected with DCGAN. DCGAN was employed to generate localized crack images. The effectiveness of the augmented dataset was validated through an image classification task. Maeda et al.[6] examined a PG-GAN to generate local synthetic images of potholes. Subsequently, larger defect-free road images were merged with the generated potholes, manually using Poisson blending, to improve object detection algorithm. Sim et al.[29] utilized an SRGAN for upscaling smaller images and implemented a semi-supervised learning approach based on an encoder-decoder architecture to enhance crack segmentation. Salaudeen et al.[30] proposed an enhanced SRGAN, trained with bounding-box pothole croppings to enlarge the dataset and improve pothole detection. The synthetic samples were hand-labelled and then, validated using YOLOv5 and EfficientDet. Chen et al.[31] and Hou et al.[32] suggested augmenting high-texture asphalt pavement using WGAN-GP architectures and expanded data for binary pavement texture image classification. Xu et al.[33] introduced a DCGAN to generate synthetic crack images from cropped drone frames and augmented the dataset for crack classification using a VGG16-based network. Analogously, Que et al.[3] developed a DCGAN to augment a crack-based dataset to enhance pavement distress classification. Song et al.[34] proposed a CycleGAN to create shadowed images of pavement cracks. The approach was validated with U-Net trained with enlarged dataset, in terms of binary segmentation. Liu et al.[35] explored a lightweight DCGAN incorporating squeeze-and-expand, multi-scale, and depth-wise modules for augmenting cracks and other defect types (pothole, patch, background). The effectiveness was validated through a classification task. Xu et al.[5] trained DCGAN to generate segmentation masks for various distress types (pothole, sealed crack, crack, distress-free). These masks and distress-free images were used to train a VAE concatenated with a discriminator to produce merged images. Zhang et al.[13] developed a DCGAN to create synthetic images, enhancing binary crack segmentation performance with a self-attention U-Net. Likewise, Pan et al.[36] introduced CrackSegAN, employing a U-Net-based generator and convolutional neural network-based discriminator for binary crack segmentation. Shim et al.[37] proposed CrackGen, a cGAN that generates synthetic images from semantic masks to enhance crack segmentation systems. However, it is limited by the manual creation of these masks for image generation. In[38], Gao et al. introduce a balanced semi-supervised GAN (BSS-GAN). The BSS-GAN combines the strengths of semi-supervised learning, which allows the model to learn from both labeled and unlabeled data, and a balanced-batch sampling technique, ensuring that each training batch includes an equal representation of all classes.

As a result of this literature review, several gaps emerge. Current generative model-based research has primarily focused on variants of GANs, enhancing segmentation and classification models, neglecting object detection. Classification methods may overlook multiple defects and lose context when applied to cropped patches. Segmentation techniques, while capable of identifying multiple defects, are computationally intensive

and require extensive manual annotation. Binary segmentation, the most explored approach, fails to provide detailed distress class information critical for effective road maintenance strategies. Moreover, non-conditional generative systems dominate, posing challenges due to labor-intensive labeling of synthetic samples limiting their applicability in road maintenance. Additionally, as mentioned earlier, while typical cracks have received significant research attention, less prevalent defects that still cause significant damage remain relatively unexplored.

To tackle these challenges, this research introduces a novel attention-guided, class-conditional Generative Adversarial Network named AsphaltGAN. This enhanced generative system incorporates multiple attention mechanisms to synthesize minor defects such as potholes, sealed cracks, diagonal, and irregular cracks from labels (minor distress types). The synthetic images are generated without requiring additional labeling efforts and are utilized to improve an object detection system. The main contributions are as follows:

1. This study designs and optimizes a novel generative model, AsphaltGAN, which is class-conditioned and enhanced with multiple attention mechanisms, capable of generating high-quality images of rare road defects that are underrepresented in the current state-of-the-art. Additionally, this model alleviates the challenges associated with image collection and addresses the costly relabeling issues commonly encountered with synthetic images in existing architectures.
2. AsphaltGAN improves upon state-of-the-art generative models by enhancing visual inspection quality, increasing image quality assessment (IQA) metrics, and strengthening class-conditioning capabilities for generating images of rare road defects.
3. Developed an end-to-end, label-efficient workflow for augmenting object detection datasets, specifically addressing object detection tasks that are less commonly studied compared to classification or segmentation. This approach demonstrates validated improvements in detection efficiency across multiple benchmark datasets.The rest of the paper is organized as follows. “Methods” introduces AsphaltGAN and the improved object detection workflow. “Datasets” briefly presents the benchmark datasets, and “Performance metrics” introduces the used IQA and object detection metrics. “Experimental results and analysis” depicts the main outcomes. The paper is concluded in “Conclusions & future scope”.

Methods

Attention-guided class-conditional generative adversarial networks

The GAN architecture consists of two primary networks: the generator and the discriminator[16]. The generator, which usually employs a noise vector sampled from a Gaussian distribution, creates synthetic images. The discriminator, acting as a binary classifier, evaluates these synthetic images alongside real ones to determine their authenticity. Both the generator and the discriminator are commonly implemented using convolutional neural networks. Key challenges in training GANs include maintaining stability, preventing mode collapse (which leads to a lack of sample diversity), managing sensitivity to hyperparameters, and addressing issues related to an unconditioned workflow. The proposed generative model will be compared with the following state-of-the-art models: cGAN[22], DCGAN[17], and WGAN-GP[19].

cGANs are an advanced form of GANs that incorporate additional information, such as specific types of asphalt distress, into both the generator and discriminator models. This conditional data helps to direct the generation process, thereby eliminating the need to label the generated images. However, while addressing one of the limitations of traditional GANs, cGANs present a new challenge known as label leakage. This occurs when information from the conditioning data unintentionally influences the generated samples, which can result in the model overfitting to the training data or limiting its ability to generalize to new, unseen data.

The WGAN is a GAN variant that uses the Wasserstein-1 distance as its loss function, resulting in more stable and efficient training. WGANs converge faster and are less sensitive to hyperparameters, making them easier to train. WGANs do not require careful balancing between the generator and discriminator, nor a meticulous architectural design, and they reduces mode collapse. WGAN employs weight clipping to maintain the Lipschitz constraint in the discriminator, preventing overfitting. In WGAN-GP, weight clipping is replaced by a Gradient Penalty (GP), further enhancing quality and stability. However, these networks lack class-conditioning, necessitating additional annotation after data augmentation.

DCGAN is an enhanced version of GAN with specific architectural improvements to address issues like noisy image generation and training instability. Key features of DCGAN include: (a) replacing pooling layers with strided convolutions in the discriminator and fractional-strided convolutions in the generator, (b) applying batch normalization in both the generator and discriminator, (c) eliminating fully connected hidden layers, (d) using Rectified Linear Unit (ReLU) activations in all generator layers except for the output, which uses hyperbolic tangent (Tanh), and (e) employing Leaky ReLU activations in all discriminator layers. DCGAN does not support class-conditioning, so additional labeling is required after data augmentation.

The architecture of AsphaltGAN, illustrated in Fig. 1, introduces a novel symmetric design incorporating distinct attention mechanisms and defect-type conditionality. The generator accepts two inputs: fake labels and a noise/latent vector. Fake labels, sampled from a uniform distribution with values between 0 and 3, correspond to four types of minor defects. These labels are converted to dense vectors via an embedding layer and reshaped to the desired size. The noise vector, generated from a Gaussian distribution, is also passed through dense and reshaping layers. Both vectors, dimensions $8 \times 8 \times 1$ for labels and $8 \times 8 \times 128$ for noise, are concatenated into a tensor of $8 \times 8 \times 129$, which then undergoes a self-attention mechanism to preserve feature integrity. The concatenated tensor passes through four identical blocks, each containing a convolutional module and spatial and channel attention mechanisms. Each convolutional module consists of four layers: 2D convolution for feature extraction, batch normalization for stability, Leaky ReLU for non-linearity, and dropout to prevent

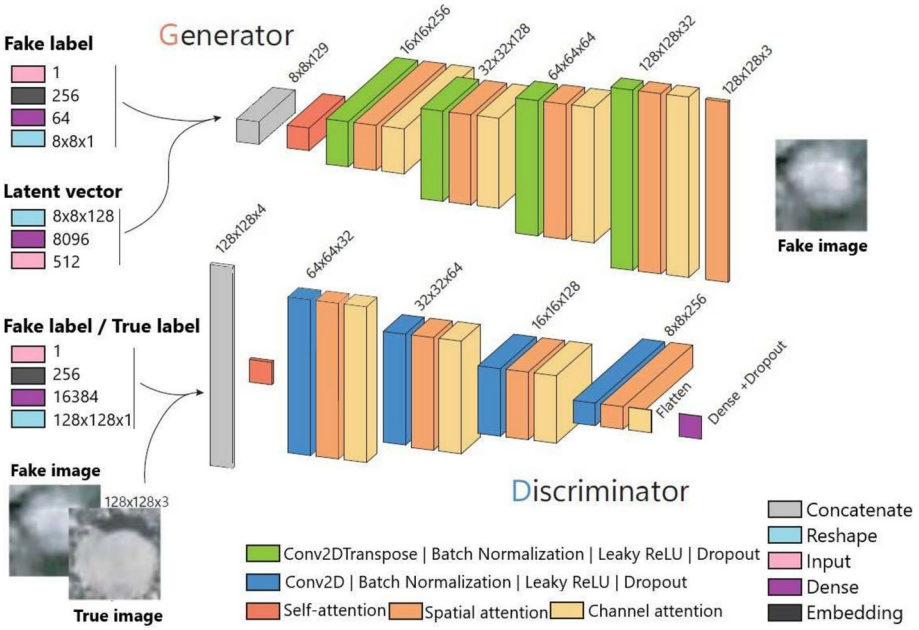

Fig. 1. Architectural design of AsphaltGAN. The symmetrical model highlights the introduction of conditionality through embedding layers to avoid relabeling synthetic samples, along with the inclusion of self-attention, channel, and spatial attention mechanisms to enhance feature learning and image generation.

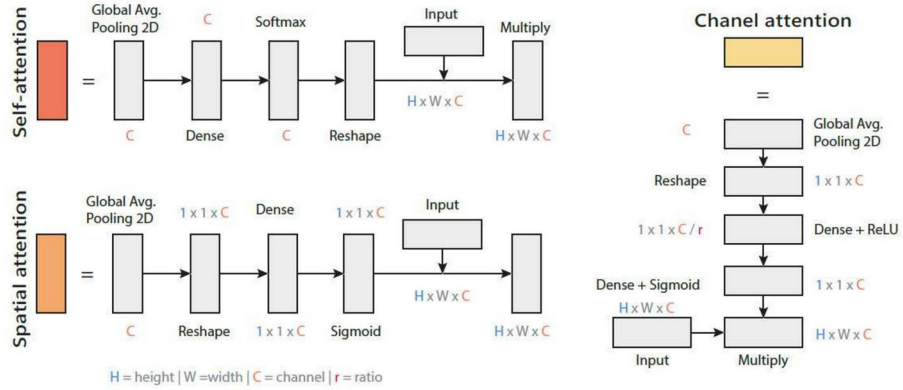

Fig. 2. Self-attention, spatial, and channel attention mechanisms included in the AsphaltGAN architecture.

overfitting. The final layer uses a similar block with Tanh activation to constrain pixel values within -1 to 1, producing a synthetic image of $128 \times 128 \times 3$.

The self-attention mechanism[39] processes the concatenated tensor by applying 2D Global Average Pooling (2D-GAP) to flatten it into a vector, capturing channel significance. This vector is transformed through dense, SoftMax, and reshaping layers to generate attention weights, which scale the tensor to enhance important features (Fig. 2). Spatial and channel attention mechanisms follow each convolutional module. Self-attention is crucial because it allows the model to weigh different parts of the input tensor, thus helping it focus on the most relevant areas of the image. This leads to better feature learning by promoting long-range dependencies across the input. In AsphaltGAN, it helps refine the features extracted from minor defects, ensuring that subtle and fine details are captured effectively. The spatial attention module uses 2D-GAP, reshaping, dense layers, and sigmoid function to generate weights that highlight significant spatial features. The spatial attention mechanism enhances feature detection by selectively focusing on regions in the image that contain crucial spatial patterns, like the distribution and shape of road defects. This localized feature enhancement improves the model's capacity to recognize important pavement details, which is critical for generating realistic synthetic frames. The channel attention module applies 2D-GAP, dense layers, and ReLU activation, followed by a sigmoid function to generate weights that emphasize important channels. Channel attention prioritizes the most relevant channels (or feature maps), enabling the model to amplify specific characteristics like texture, depth, or edges, which are essential

	Generator	Discriminator
BCE-GP	$BCE(y^r, y_p^r) + BCE(y^f, y_p^f) + GP$	$BCE(y^r, y_p^f)$
MSE-GP	$MSE(y^r, y_p^r) + MSE(y^f, y_p^f) + GP$	$MSE(y^r, y_p^f)$
H-GP	$H(y^r, y_p^r) + H(y^f, y_p^f) + GP$	$H(y^r, y_p^f)$

Table 1. Loss functions: binary cross entropy with GP (BCE-GP), mean squared error with GP (MSE-GP), and hinge with GP (H-GP). The analytical expressions are in[16,40,41], respectively.

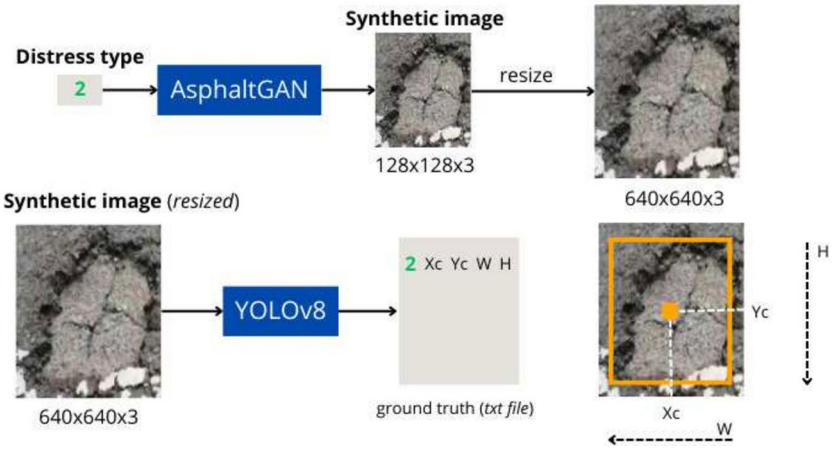

Fig. 3. Improved end-to-end object detection methodology. The diagram highlights two processes: (1) training AsphaltGAN using labeled images with a single defect extracted from multi-defect annotations, and (2) resizing synthetic images from AsphaltGAN and including them in the training set with programmatically generated ground truth for YOLOv8 training.

in identifying fine-grained distress patterns. By integrating this with spatial attention, AsphaltGAN boosts its ability to enhance both spatial structure and channel-specific features progressively throughout the network.

The discriminator has four inputs: real/fake labels and real/fake images. Labels are embedded, passed through a dense layer, and reshaped to $128 \times 128 \times 1$, then concatenated with images into a tensor of $128 \times 128 \times 4$. The discriminator mirrors the generator's architecture with self-attention and convolutional blocks, but it reduces spatial dimensions and increases color dimensions using transposed convolutions. The output tensor is flattened and passed through a dense layer with dropout, yielding values for each input image.

Loss functions

Due to the complexity of GAN training, various loss functions will be used to assess the proposed model's performance (Table 1). Each discriminator's loss function will include a GP to penalize large gradients, reducing overfitting and promoting stability. The optimal loss function will be selected based on empirical results. The labels y^r, y^f, y_p^r, and y_p^f denote real, fake, predicted real, and predicted fake images, respectively. These labels indicate image authenticity rather than defect classes, with zeros for fake images and ones for real ones.

Improved object detection pipeline

The objective of this study is to propose an end-to-end methodology to enhance deep learning-based object detection systems for identifying minor pavement defects. This includes addressing issues such as the system's ability to handle imbalanced defect classes and lower detection rates. Additionally, the process should be fully automated, eliminating the need for human intervention, such as the relabeling of synthetic images, to ensure a truly end-to-end solution for engineering applications.

The pipeline of this study, depicted in Fig. 3, involved labeling an asphalt pavement distress image dataset (Mosquitonet[42]) for enhancing an object detection system. Observations showed that less common defects (potholes, sealed cracks, diagonal cracks, and irregular cracks) had lower detection rates, these defects were cropped and resized to 128x128x3 from the bounding box annotations. The pictures extracted from the dataset exhibit multiple defects of various types per image. Minor defects identified through annotated labels in an object detection framework, were programmatically cropped to produce smaller images containing a single defect instance per image. These images and their labels trained AsphaltGAN to generate synthetic images of minority defects. The synthetic images were resized to match the original dataset dimensions ($640 \times 640 \times 3$), with bounding boxes centered and scaled to 90% of the original size. This process aimed to validate improved detection performance and model robustness across various scales. The object detection system was trained

Dataset (defect)	RDD2022 (pothole)	APDD (diagonal crack)	CrackSC (irregular crack)	CrackDLHY (sealed crack)
Train	188	138	42	400
Test	47	34	11	101

Table 2. Distribution of train/test images by object detection benchmark dataset and minor distress class.

	Sealed crack	Diagonal crack	Irregular crack	Pothole
Train	350	394	232	596
Test	88	98	59	149

Table 3. Distribution of less-represented superficial road defects in the Mosquitonet subdataset.

using both original and synthetic images to enhance detection automatically without human involvement, thereby addressing the limitation of needing more image data.

Datasets

The dataset used to train and validate the various generative models is an open-source image repository called Mosquitonet[42]. To validate our data augmentation system's robustness, we selected various public object detection datasets that include minor asphalt pavement defects: Road Damage Dataset 2022 (RDD2022)[43], CrackDLHY[44], APDD[45], and CrackSC[28]. This approach avoids potential data contamination from Mosquitonet and assesses the generalization of our pre-processing technique. Since not all datasets cover every minority defect, the evaluation will focus on detection performance for the specific augmented defects. Each benchmark dataset will use an 80/20 train/test split, with class distributions shown in Table 2.

Mosquitonet

It is a top-down-view benchmark dataset collected with low-cost vehicle-mounted camera. There are 7099 images of 13 asphalt pavement defects of $640 \times 640 \times 3$, annotated by pavement experts in several object detection formats. For our study, leveraging the annotations (bounding boxes and categories), patches of the minor defects (sealed, diagonal, irregular crack, and pothole) along with the distress type were programmatically extracted, resized to $128 \times 128 \times 3$ (average patch dimensions), and normalized. Thus, a sub-dataset of 1966 instances (local defect images) with 4 classes was constructed, with a train split of 80% and a 20% test split. The distribution per distress type is shown in Table 3.

RDD2022

RDD2022 comprises 47420 road images with 55000 instances object detection annotations. It contains four road damages: longitudinal, transverse, alligator crack, and pothole. The sub-dataset that most closely approximates the top-down view configuration (selected to prevent network confusion, e.g., diagonal cracks that may appear longitudinal in a wide-view perspective) is China-Motorbike. It was collected with motorbike-mounted smartphone camera moving at an average speed of 30 km/h. This sub-dataset contains 1977 images of $512 \times 512 \times 3$.

CrackDLHY

CrackDLHY contains a box-based image dataset with four distress types: alligator, sealed, transverse and longitudinal crack. The collection combines manual smart phone photography and vehicle-mounted camera. The frames are captured with different light intensities and shooting angles. The image size is $1280 \times 960 \times 3$.

APDD

APDD includes 3150 images of $512 \times 512 \times 3$ with alligator, block, longitudinal, diagonal, transverse crack, repair, and pothole.

CrackSC

CrackSC is annotated for segmentation, but it contains multiple isolated irregular cracks (without defined geometry). Hence, they have been labelled in object detection format. The images were captured using a smartphone on local roads, where heavy shadows and artefacts (e.g., leaves) were present. There are 53 images with dimensions of $320 \times 480 \times 3$.

Performance metrics

IQA metrics

The image quality of the sampled images generated by the proposed generative architectures was assessed both visually and quantitatively using standard IQA metrics. These metrics, including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Fréchet Inception Distance (FID), were used in the optimization of AsphaltGAN and to compare its optimum design with state-of-the-art generative models.

The IQA metrics were computed for both the test set and an equivalent number of images produced by the corresponding generative model.

PSNR measures image quality by comparing the maximum possible pixel value to the MSE between the real and the synthetic image. A higher PSNR indicates a better quality. While PSNR is commonly used, it may not fully capture perceptual differences. SSIM evaluates image quality based on luminance, contrast, and structure providing a more perceptual assessment. The closer the SSIM is to one, the better the image quality. FID measures the dissimilarity between feature distributions of real and generated images using a pretrained neural network, with lower FID indicating higher similarity and better quality. The analytical expressions for PSNR, SSIM, and FID are found in[46].

Object detection metrics

To validate object detection model improvements, several metrics are employed. Precision measures the ratio of correctly identified relevant instances to the total predicted positives, reflecting accuracy but potentially missing some true positives. Recall indicates the ratio of correctly identified instances to all actual positives, highlighting the model's ability to capture relevant instances while risking more false positives. Mean Average Precision (mAP) averages the area under the precision-recall curve across confidence thresholds, with higher scores denoting better performance. mAP_{50} evaluates at an Intersection over the Union (IoU) threshold of 0.5, while mAP_{50-95} extends the assessment to IoU thresholds from 0.5 to 0.95, offering insights into performance across various object sizes and poses. The mathematical formulas are detailed in[47].

Experimental results and analysis

In the results section, AsphaltGAN is first analyzed in terms of various loss functions and hyperparameter tuning. The optimal configuration is then compared with state-of-the-art generative models. Finally, results are presented for the object detection models following the application of synthetic data augmentation.

For AsphaltGAN, the LeakyReLU activation functions use a value of $\alpha = 0.2$, and dropout rates of 0.3 and 0.4 are applied in the generator (G) and discriminator (D) layers, respectively. Hyperparameter tuning included sensitivity analysis on the balance between G and D training steps per epoch (k) and learning rate schedules (linear decay, cosine annealing, and constant). The designs of WGAN-GP, DCGAN, and cGAN were based on convolutional layers, aiming to maintain a similar complexity (number of parameters) to AsphaltGAN. Common hyperparameters for all generative models are: learning rates of 1e−4 for both G and D, a batch size of 128, Adam optimizers, a latent dimension of 512, a GP weight of 10, and 3000 training epochs. YOLOv8[48] was used as the validation model, trained on real and synthetic images. The code is implemented in Python with TensorFlow 2.8.0 for generative models and OpenCV 4.7.0 for image processing. YOLOv8 is implemented in PyTorch 2.0.1. The setup includes a Dell Alienware Aurora R13 with 64 GiB of memory, CUDA 11.4, and an NVIDIA GeForce RTX 3080 Ti GPU, with development conducted in Visual Studio Code.

Optimizing AsphaltGAN: loss functions & hyperparameters

The analysis starts with BCE-GP (Fig. 4), where early training (up to epoch 500) shows a slight rise in both train and test losses due to the model's adaptation process. Following this, the test loss slightly increases while train loss decreases, indicating that D might be overfitting on known images but performing well on training data. D stabilizes with minor fluctuations around 2.36 ± 0.01 (test loss) and 0.70 ± 0.02 (train loss) from epoch 2500 onward. Meanwhile, G consistently improves, stabilizing at 1.43 ± 0.01 for both losses. The synthetic images generated are realistic and diverse (Fig. 5). In contrast, MSE-GP exhibits a faster learning rate but suffers from artifacts and noise after epoch 2250, leading to instability in loss functions. G's performance degrades with increased artifacts. Figure 5 shows synthetic images before instability, demonstrating the model's proficiency in creating diverse and realistic potholes and sealed cracks, though it struggles with granular textures and introduces artifacts in diagonal and irregular cracks. For H-GP, D shows low training loss but high test loss fluctuations, indicating overfitting and poor generalization. G's loss plateaus, reflecting mode collapse and limited texture diversity. H-GP fails to capture detailed defect features effectively, making it less suitable. W-GP loss faced convergence issues and was thus excluded. IQA metrics (Table 4) further show H-GP's lower PSNR and SSIM, and higher FID compared to BCE-GP and MSE-GP, confirming the trends observed in loss function dynamics. BCE-GP achieves the best IQA metrics.

In GAN training, D is often trained more than G to ensure stability, prevent mode collapse, and improve gradient flow. The following sub-analysis (Fig. 6) explores various additional epochs (k) for D in AsphaltGAN, showing that $k = 5$ achieves the best IQA metrics, albeit at the cost of higher computational demand. Lower k values did not consistently improve PSNR, SSIM, or FID, blueindicating that longer training of D is necessary for optimal performance. Monitoring loss functions revealed that for $k = 1$, both D and G losses stabilized by epoch 2000, whereas for $k = 3$, instability occurred from epoch 1200 onward. This suggests that higher k values contribute to a more stable convergence but require careful tuning to avoid instability. Additionally, different learning rate (LR) schedules were tested (Fig. 6), with linear decay providing the best FID results, while the constant LR produced the highest SSIM. Cosine annealing exhibited instability throughout the training process, especially at later epochs. Thus, the optimal AsphaltGAN configuration is BCE-GP/BCE, $k = 5$, and a constant LR of $1e - 4$, which demonstrated significant improvements in IQA metrics, training stability, and convergence.

Figure 7 presents synthetic images generated by AsphaltGAN with and without the inclusion of attention mechanisms, illustrating their positive impact as outlined in the previous section. For instances of diagonal and irregular cracks, the addition of attention mechanisms not only yields more precisely defined cracks but also achieves a more granular, textured background that resembles the coarse texture of asphalt pavements due to aggregate presence. In the case of potholes, where the defect occupies a substantial portion of the image, the attention mechanisms enhance edge resolution and create a depth effect by projecting shadows, improving the

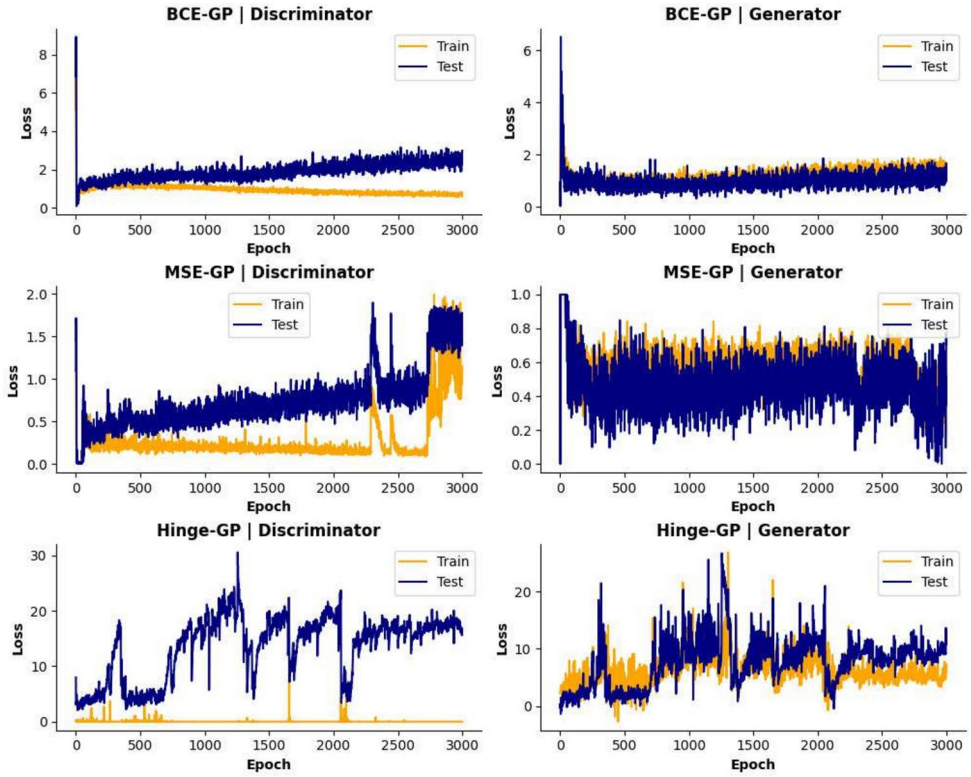

Fig. 4. Loss functions for the generator and discriminator of AsphaltGAN.

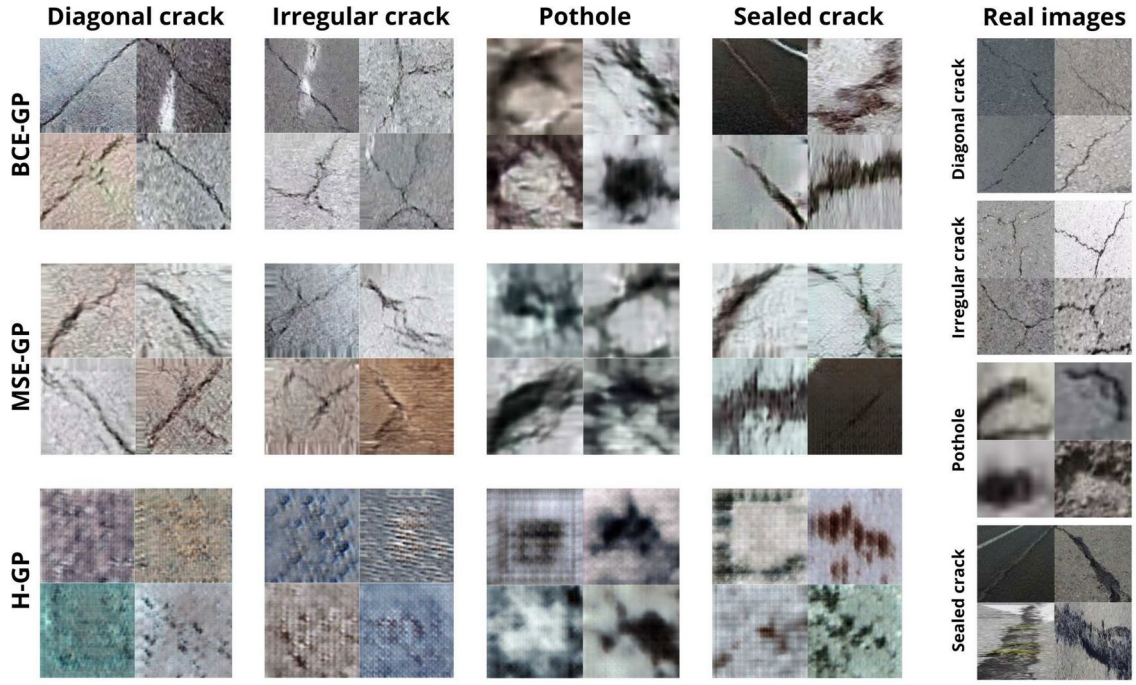

Fig. 5. Synthetic minor asphalt road images generated by AsphaltGAN using various loss functions.

realism of the synthetic defects. For sealed cracks, the model without attention mechanisms struggles to generate large-scale sealed cracks, often producing backgrounds with minimal texture and occasionally introducing random artifacts along the edges. However, with attention mechanisms, AsphaltGAN successfully generates extensive sealed cracks, enhances background texture, and even introduces road markings in certain images.

	PSNR	SSIM	FID
BCE-GP	25.93 ± 0.03	0.81 ± 0.01	9.58 ± 0.04
MSE-GP	21.83 ± 0.03	0.56 ± 0.02	11.43 ± 0.13
H-GP	10.96 ± 0.04	0.11 ± 0.01	49.02 ± 0.48

Table 4. IQA metrics for AsphaltGAN using different loss functions.

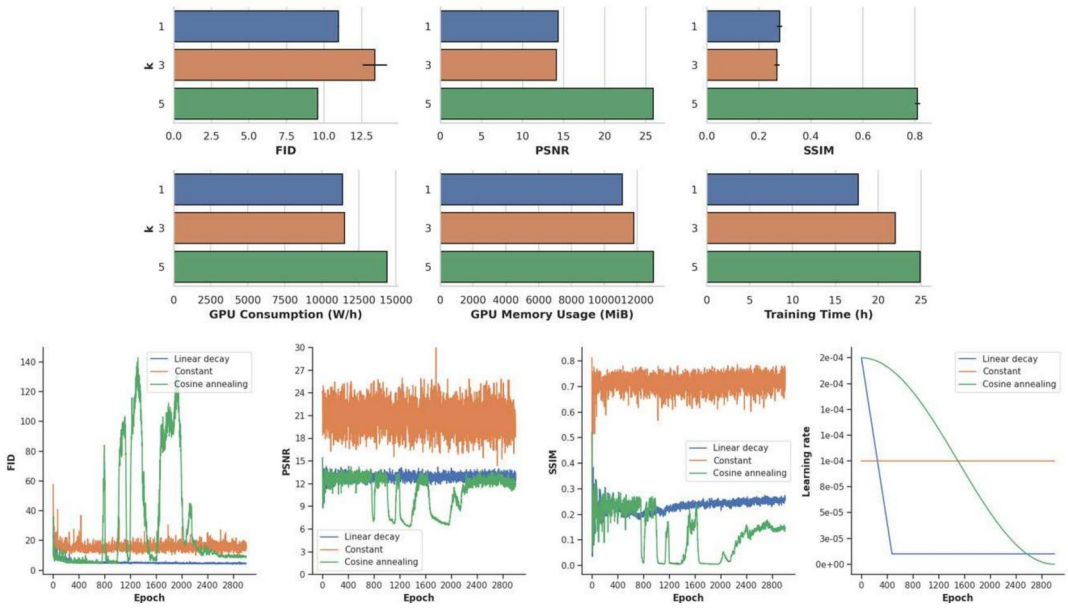

Fig. 6. (Top) Fine-tuning additional training steps of D in terms of computational cost and IQA metrics. (Bottom) IQA metric curves for different learning rate schedules.

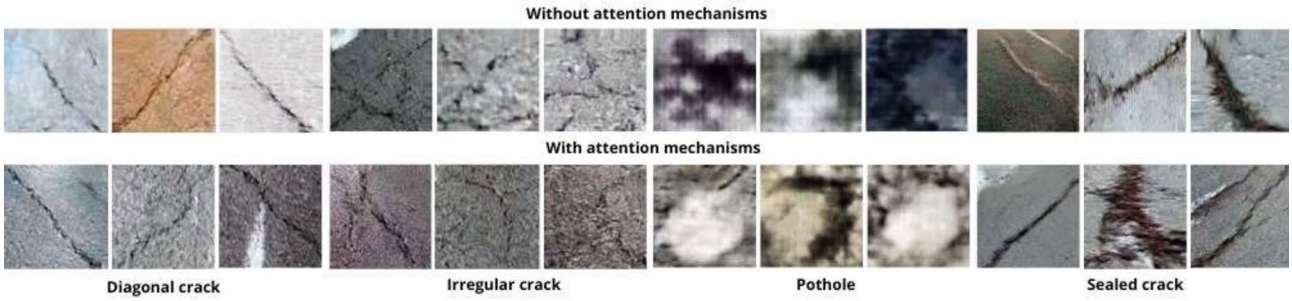

Fig. 7. Synthetic images generated by AsphaltGAN under optimal settings for each type of minor defect, with and without the inclusion of attention mechanisms, demonstrating the positive impact of attention mechanisms on the results.

These attention mechanisms enable AsphaltGAN to progressively refine both spatial and channel-specific features, yielding synthetic images that capture critical pavement details with enhanced realism.

Comparison with state-of-the-art generative architectures

cGAN and AsphaltGAN exhibit similar loss function curves, demonstrating clear convergence and high learning stability (Fig. 8). WGAN-GP, despite higher oscillations due to its unbounded nature, also shows convergence and stability. In contrast, DCGAN's loss functions indicate inadequate convergence, with discriminator loss hovering around zero and oscillating significantly, suggesting overfitting and hindering generator performance. Oscillating generator losses reinforce this observation, highlighting the generator's struggle to deceive the discriminator. DCGAN's conspicuous oscillations in the FID curves highlight poor convergence, while WGAN-GP and AsphaltGAN demonstrate smoother FID curves and lower average values, indicating better convergence

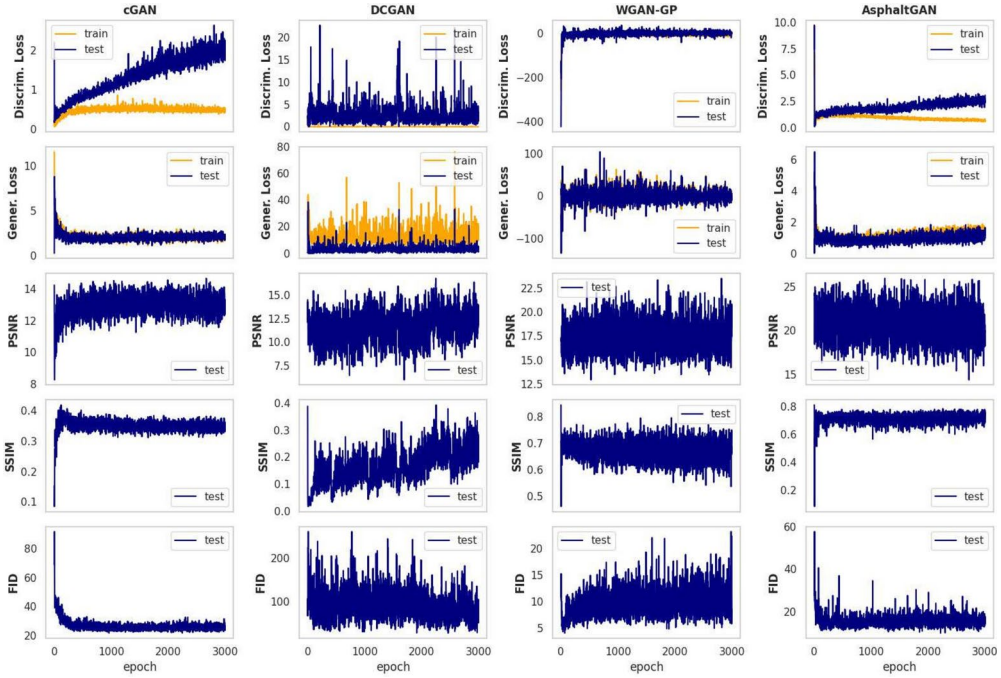

Fig. 8. Comparison of loss and image quality metrics for generative architectures using the test split of the public Mosquitonet dataset.

Generative model	PSNR	SSIM	FID
DCGAN	12.20 ± 0.04	0.22 ± 0.01	28.47 ± 0.60
cGAN	12.87 ± 0.02	0.36 ± 0.01	26.70 ± 0.09
WGAN-GP	17.49 ± 0.03	0.67 ± 0.11	10.16 ± 0.07
AsphaltGAN	25.93 ± 0.03	0.81 ± 0.01	9.58 ± 0.04

Table 5. IQA metrics for several generative architectures.

and image quality. AsphaltGAN achieves superior PSNR and SSIM values, with WGAN-GP and AsphaltGAN excelling in FID (Table 5).

Figure 9 visualizes synthetic images generated during training. In Fig. 10, synthetic frames are presented by defect type for each of the pre-trained generative models. DCGAN only produces noisy pothole-type defects, consistent with its poor loss function and image quality metrics. The remaining defects are poorly defined and appear unrealistic with the results aligning with the previously analyzed lack of convergence in the loss functions. Generative models first learn to generate potholes, the most common asphalt defect. Learning order varies by model, with sealed cracks and diagonal cracks being learned differently. Irregular or undefined geometries are the most challenging, likely due to their geometric complexity and underrepresentation in the training dataset compared to other minor defects. WGAN-GP performs well on most defects but struggles with textures of irregular and wide-perspective diagonal cracks. In fact, irregular cracks exhibit poorly defined geometries and struggle to connect the various branches of the crack network. cGAN faces challenges with large sealed cracks and those in low-light conditions. Additionally, for diagonal and irregular cracks, streaked artifacts often appear along the edges in many images, confirming the poorer results in image quality metrics. AsphaltGAN generates a diverse range of defects but struggles with irregular cracks. The generative architectures have comparable complexity and similar training and inference times, differing only in convergence time, with WGAN-GP being slightly faster. In contrast, AsphaltGAN's conditionality advantage simplifies synthetic image labeling, reducing manual effort.

Additionally, a human visual inspection experiment was conducted, generating 100 synthetic images per model to evaluate the effectiveness of conditionality. Conditional models used defect-specific vectors, while non-conditional models generated 100 general images. Ideally, each defect type should have 25 images. DCGAN failed to produce diagonal cracks, whereas other models generated around 15 images of this type. AsphaltGAN performed best for irregular cracks, while WGAN-GP showed a bias towards potholes, producing over 30 images of this defect. AsphaltGAN achieved more balanced results overall. For sealed cracks, WGAN-GP performed best, followed by AsphaltGAN. AsphaltGAN's conditionality reduces the need for manual labeling. In summary, DCGAN has convergence and stability issues, limiting it to producing noisy pothole defects. Conversely, WGAN-GP, AsphaltGAN, and cGAN demonstrate sound convergence and stability, with AsphaltGAN achieving the best

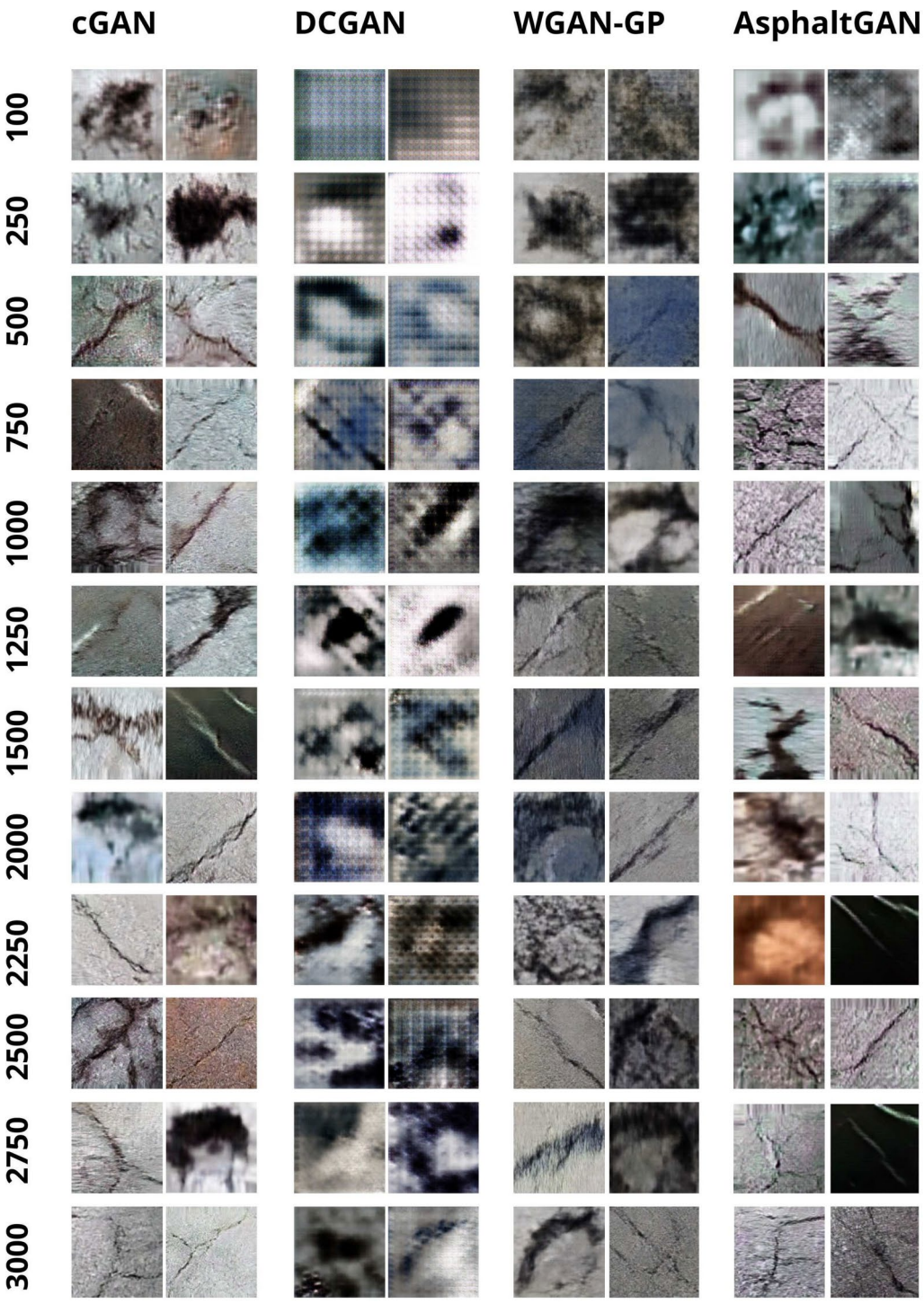

Fig. 9. Synthetic images generated by various algorithms during the training stages.

IQA metrics and consistently generating all defect types. AsphaltGAN's conditionality is advantageous, making it the most suitable model for generating realistic asphalt defect images. The proposed generator will be used to enhance public datasets, improving YOLOv8 detection efficiency.

Improved object detection system with synthetic images

To determine the number of synthetic images to add per dataset, a histogram categorized by defect type and the mean value was displayed. The minority defects, targeted for augmentation through AsphaltGAN-generated synthetic samples, were adjusted to approach the mean value to promote class balance. Object detection metrics, detailed in Table 6, are provided before and after augmentation. It is important to note that only results for the minority defects, the focus of this study, are presented. Increasing the representation of minority defects

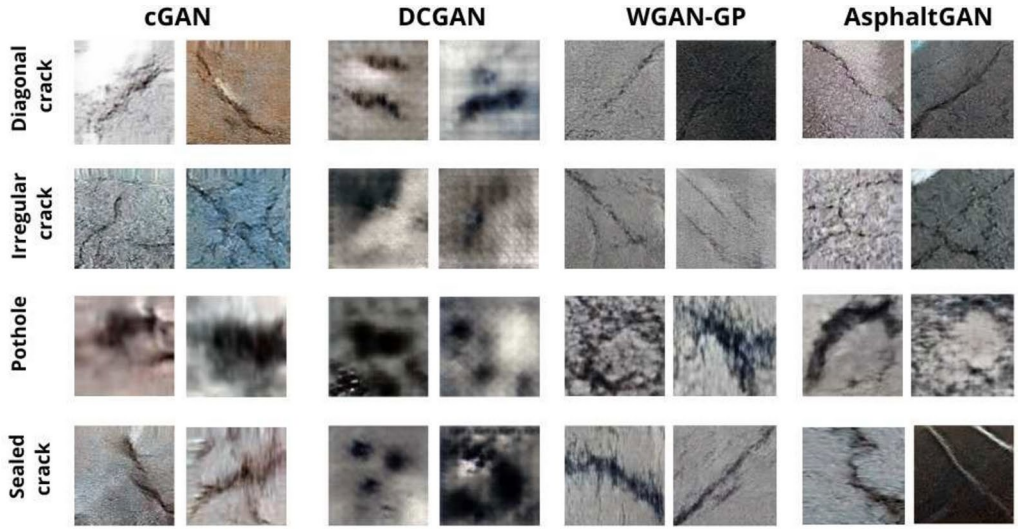

Fig. 10. Fake frames produced by multiple pre-trained generative architectures.

Dataset	Distress	Configuration	Precision	Recall	mAP_{50}	mAP_{50-95}	mAP_{50}^{all}
RD2022	Pothole	YOLOv8	0.489	0.167	0.201	0.102	0.524
		YOLOv8+AsphaltGAN	0.837	0.267	0.304	0.214	0.697
CrackDLHY	Sealed crack	YOLOv8	0.463	0.640	0.494	0.269	0.418
		YOLOv8+AsphaltGAN	0.497	0.651	0.506	0.288	0.431
APDD	Diagonal crack	YOLOv8	0.100	0.258	0.100	0.029	0.353
		YOLOv8+AsphaltGAN	0.380	0.310	0.293	0.183	0.518
CrackSC	Irregular crack	YOLOv8	0.367	0.636	0.479	0.240	0.479
		YOLOv8+AsphaltGAN	0.702	0.727	0.724	0.342	0.724

Table 6. Object detection results with several benchmark asphalt defect.

to enhance their detection metrics may impact the metrics for majority defects, which are not analyzed here. Therefore, the mAP_{50}^{all} metric, reflecting the mAP_{50} averaged across all defect types, is included. The results show a significant improvement in precision, indicating fewer false-positive predictions by YOLOv8. Recall also improved, showing more true positive detections. Consequently, there is a notable increase in mAP_{50} and mAP_{50-95} metrics, contributing to an average enhancement across all defect categories. Importantly, the improved detection metrics for minority defects do not negatively impact the performance on majority defects. This is evidenced by the improvement in mAP_{50}^{all} across all datasets: RDD2022 (↑33.0%), CrackDLHY (↑3.1%), APDD (↑46.7%), and CrackSC (↑51.2%).

The improved defect detection system using AsphaltGAN images has certain limitations (Fig. 11). First, the enhancement in detecting sealed cracks is minimal. This is likely due to the CrackDLHY dataset being captured with a perpendicular angle to the pavement, while the Mosquitonet dataset, used for training AsphaltGAN to generate images, incorporates a slight incidence angle to capture all lane defects. As a result, the model struggles to improve detection of sealed cracks, which often appear with a diagonal geometry due to perspective. Another limitation arises in detecting minor defects with highly variable dimensions. During AsphaltGAN training, an average resize was applied to adapt to convolutional operations, which may not affect image generation significantly, as smaller defects typically occupy a reduced portion of the image. However, sealed cracks, which can span large areas, are unrealistically scaled down due to this resizing, as their average size is skewed by smaller defects like potholes or irregular cracks. This highlights potential challenges when transferring AsphaltGAN to datasets with different image perspectives or object size distributions, suggesting that future improvements should consider dataset-specific geometry and defect size variability to further enhance detection performance.

Conclusions & future scope

This study presents AsphaltGAN, a class-conditional Generative Adversarial Network incorporating self-attention, spatial, and channel attention mechanisms, designed to generate synthetic images of infrequent asphalt pavement defects. These synthetic images augment public datasets, enhancing the deep learning-based object detection performance of YOLOv8. The main conclusions are as follows:

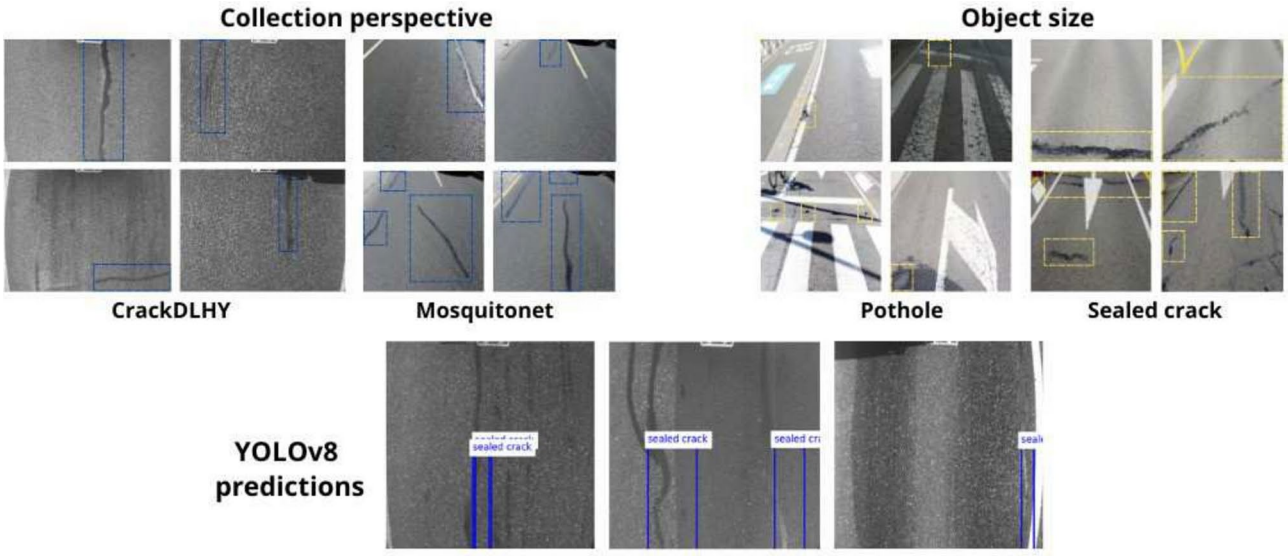

Fig. 11. Presentation of limitations identified in AsphaltGAN. (Top-left) The difference in perspective between the image collection used to train AsphaltGAN and those used for YOLOv8 may impact detection improvements. (Top-right) Intra- and inter-class size differences may affect generation due to rescaling, potentially impacting the generated images. (Bottom) Incorrect detections from YOLOv8 enhanced with AsphaltGAN from CrackDLHY dataset.

- AsphaltGAN was trained and tested using various loss functions: BCE-GP, MSE-GP, and H-GP, all with gradient penalty. The BCE-GP configuration demonstrated the most stable convergence and superior IQA metrics.
- Hyperparameter tuning identified the optimal configuration with an additional five discriminator training epochs and a constant learning rate of 1×10^{-4}.
- AsphaltGAN outperforms state-of-the-art models, including WGAN-GP, DCGAN, and cGAN, with superior image quality metrics on the public Mosquitonet dataset (PSNR: 25.93, SSIM: 0.81, FID: 9.58) and effectively generates diverse minority defects without the need for synthetic image labeling.
- The YOLOv8-AsphaltGAN architecture significantly improved the detection of minority defects in public datasets (RDD2022, CrackDLHY, APDD, CrackSC), reducing the need for manual data collection and labeling. Future research will focus on addressing the identified limitations of AsphaltGAN, with particular emphasis on designing generative models capable of blending defect-free images with cropped images of minor defects, extracted from bounding boxes, to avoid rescaling issues with defects of varying sizes. Additionally, the implementation of semantic synthesis generative architectures will be explored to overcome the challenges posed by the angle of incidence between datasets, ensuring better control over the geometries of minor defects.

Data availability

The datasets can be accessed via the following URLs: Mosquitonet at https://repositorio.unican.es/xmlui/handle/10902/26615, RDD2022 at https://github.com/sekilab/RoadDamageDetector, CrackDLHY at https://github.com/juhuyan/CrackDataset_DL_HY, APDD at https://universe.roboflow.com/pavement-distresses-detection/asphalt_distress_detection, and CrackSC https://github.com/KangchengLiu/Crack-Detection-and-Segmentation-Dataset-for-UAV-Inspection.

Received: 9 August 2024; Accepted: 15 November 2024
Published online: 21 November 2024

References

1. Ai, D., Jiang, G., Lam, S.-K., He, P. & Li, C. Computer vision framework for crack detection of civil infrastructure-A review. *Eng. Appl. Artif. Intell.* **117**, 105478. https://doi.org/10.1016/j.engappai.2022.105478 (2023).
2. Koch, C., Georgieva, K., Kasireddy, V., Akinci, B. & Fieguth, P. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Adv. Eng. Inform.* **29**, 196–210. https://doi.org/10.1016/j.aei.2015.01.008 (2015) (infrastructure computer vision).
3. Que, Y. et al. Automatic classification of asphalt pavement cracks using a novel integrated generative adversarial networks and improved vgg model. *Eng. Struct.* **277**, 115406. https://doi.org/10.1016/j.engstruct.2022.115406 (2023).
4. Cano-Ortiz, S., Pascual-Muñoz, P. & Castro-Fresno, D. Machine learning algorithms for monitoring pavement performance. *Autom. Construct.* **139**, 104309. https://doi.org/10.1016/j.autcon.2022.104309 (2022).
5. Xu, Z. et al. Enhancing pavement distress detection using a morphological constraints-based data augmentation method. *Coatings* **13**, 764. https://doi.org/10.3390/coatings13040764 (2023).
6. Maeda, H., Kashiyama, T., Sekimoto, Y., Seto, T. & Omata, H. Generative adversarial network for road damage detection. *Comput. -Aided Civ. Infrastruct. Eng.* **36**, 47–60. https://doi.org/10.1111/mice.12561 (2020).
7. Shang, J. et al. Automatic pixel-level pavement sealed crack detection using multi-fusion u-net network. *Measurement* **208**, 112475. https://doi.org/10.1016/j.measurement.2023.112475 (2023).

8. El Hakea, A. H. & Fakhr, M. W. Recent computer vision applications for pavement distress and condition assessment. *Autom. Construct.* **146**, 104664. https://doi.org/10.1016/j.autcon.2022.104664 (2023).
9. Li, P. et al. Our-net: A multi-frequency network with octave max unpooling and octave convolution residual block for pavement crack segmentation. *IEEE Trans. Intell. Transport. Syst.* **25**, 13833–13848. https://doi.org/10.1109/TITS.2024.3405995 (2024).
10. Zhu, G. et al. A lightweight encoder-decoder network for automatic pavement crack detection. *Comput.-Aided Civ. Infrastruct. Eng.* **39**, 1743–1765. https://doi.org/10.1111/mice.13103 (2023).
11. Zhong, J. et al. A deeper generative adversarial network for grooved cement concrete pavement crack detection. *Eng. Appl. Artif. Intell.* **119**, 105808. https://doi.org/10.1016/j.engappai.2022.105808 (2023).
12. Deng, J., Lu, Y. & Lee, V. C. A hybrid lightweight encoder-decoder network for automatic bridge crack assessment with real-world interference. *Measurement* **216**, 112892. https://doi.org/10.1016/j.measurement.2023.112892 (2023).
13. Zhang, T., Wang, D., Mullins, A. & Lu, Y. Integrated apc-gan and attunet framework for automated pavement crack pixel-level segmentation: A new solution to small training datasets. *Trans. Intell. Transport. Sys.* **24**, 4474–4481. https://doi.org/10.1109/TITS.2023.3236247 (2023).
14. He, X. et al. A survey of defect detection applications based on generative adversarial networks. *IEEE Access* **10**, 113493–113512. https://doi.org/10.1109/ACCESS.2022.3217227 (2022).
15. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. arXiv:1312.6114 (2022).
16. Goodfellow, I. J. *et al.* Generative adversarial networks. arXiv:1406.2661 (2014).
17. Radford, A., Metz, L. & Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434 (2016).
18. Arjovsky, M., Chintala, S. & Bottou, L. Wasserstein GAN. arXiv:1701.07875 (2017).
19. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. & Courville, A. Improved training of Wasserstein GANs. arXiv:1704.00028 (2017).
20. Karras, T., Aila, T., Laine, S. & Lehtinen, J. Progressive growing of GANs for improved quality, stability, and variation. arXiv:1710.10196 (2018).
21. Ledig, C. *et al.* Photo-realistic single image super-resolution using a generative adversarial network. arXiv:1609.04802 (2017).
22. Mirza, M. & Osindero, S. Conditional generative adversarial nets. https://doi.org/10.48550/ARXIV.1411.1784 (2014).
23. Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv:1703.10593 (2020).
24. Mazzini, D., Napoletano, P., Piccoli, F. & Schettini, R. A novel approach to data augmentation for pavement distress segmentation. *Comput. Indus.* **121**, 103225. https://doi.org/10.1016/j.compind.2020.103225 (2020).
25. Mei, Q. & Gül, M. A cost effective solution for pavement crack inspection using cameras and deep neural networks. *Construct. Build. Mater.* **256**, 119397. https://doi.org/10.1016/j.conbuildmat.2020.119397 (2020).
26. Zhang, K., Zhang, Y. & Cheng, H.-D. Crackgan: Pavement crack detection using partially accurate ground truths based on generative adversarial learning. *IEEE Trans. Intell. Transport. Syst.* **22**, 1306–1319. https://doi.org/10.1109/TITS.2020.2990703 (2021).
27. Gao, Y., Kong, B. & Mosalam, K. M. Deep leaf-bootstrapping generative adversarial network for structural image data augmentation. *Comput.-Aided Civ. Infrastruct. Eng.* **34**, 755–773. https://doi.org/10.1111/mice.12458 (2019).
28. Pei, L. et al. Virtual generation of pavement crack images based on improved deep convolutional generative adversarial network. *Eng. Appl. Artif. Intell.* **104**, 104376. https://doi.org/10.1016/j.engappai.2021.104376 (2021).
29. Shim, S., Kim, J., Lee, S.-W. & Cho, G.-C. Road damage detection using super-resolution and semi-supervised learning with generative adversarial network. *Autom. Construct.* **135**, 104139. https://doi.org/10.1016/j.autcon.2022.104139 (2022).
30. Salaudeen, H. & Çelebi, E. Pothole detection using image enhancement GAN and object detection network. *Electronics* **11**. https://doi.org/10.3390/electronics11121882 (2022).
31. Chen, N. et al. Data augmentation and intelligent recognition in pavement texture using a deep learning. *IEEE Trans. Intell. Transport. Syst.* **23**, 25427–25436. https://doi.org/10.1109/TITS.2022.3140586 (2022).
32. Hou, Y. et al. A deep learning method for pavement crack identification based on limited field images. *IEEE Trans. Intell. Transport. Syst.* **23**, 22156–22165. https://doi.org/10.1109/TITS.2022.3160524 (2022).
33. Xu, B. & Liu, C. Pavement crack detection algorithm based on generative adversarial network and convolutional neural network under small samples. *Measurement* **196**, 111219. https://doi.org/10.1016/j.measurement.2022.111219 (2022).
34. Song, J., Li, P., Fang, Q., Xia, H. & Guo, R. Data augmentation by an additional self-supervised cyclegan-based for shadowed pavement detection. *Sustainability* **14**. https://doi.org/10.3390/su142114304 (2022).
35. Liu, Z. et al. Automatic intelligent recognition of pavement distresses with limited dataset using generative adversarial networks. *Autom. Construct.* **146**, 104674. https://doi.org/10.1016/j.autcon.2022.104674 (2023).
36. Pan, Z., Lau, S. L., Yang, X., Guo, N. & Wang, X. Automatic pavement crack segmentation using a generative adversarial network (GAN)-based convolutional neural network. *Results Eng.* **19**, 101267. https://doi.org/10.1016/j.rineng.2023.101267 (2023).
37. Shim, S. Self-training approach for crack detection using synthesized crack images based on conditional generative adversarial network. *Comput.-Aid. Civ. Infrastruct. Eng.* **39**, 1019–1041. https://doi.org/10.1111/mice.13119 (2023).
38. Gao, Y., Zhai, P. & Mosalam, K. M. Balanced semisupervised generative adversarial network for damage assessment from low-data imbalanced-class regime. *Comput.-Aided Civ. Infrastruct. Eng.* **36**, 1094–1113. https://doi.org/10.1111/mice.12741 (2021).
39. Vaswani, A. *et al.* Attention is all you need. https://doi.org/10.48550/ARXIV.1706.03762 (2017).
40. Mao, X. *et al.* Least squares generative adversarial networks. https://doi.org/10.48550/ARXIV.1611.04076 (2016).
41. Kavalerov, I., Czaja, W. & Chellappa, R. A multi-class hinge loss for conditional GANs. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 1289–1298. https://doi.org/10.1109/WACV48630.2021.00133 (2021).
42. Cano-Ortiz, S., Lloret Iglesias, L., Martinez Ruiz del Árbol, P., Lastra-González, P. & Castro-Fresno, D. An end-to-end computer vision system based on deep learning for pavement distress detection and quantification. *Construct. Build. Mater.* **416**, 135036 https://doi.org/10.1016/j.conbuildmat.2024.135036 (2024).
43. Tayo, C. O., Linsangan, N. B. & Pellegrino, R. V. Portable crack width calculation of concrete road pavement using machine vision. In *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*. https://doi.org/10.1109/hnicem48295.2019.9072731 (IEEE, 2019).
44. He, X. et al. A survey of defect detection applications based on generative adversarial networks. *IEEE Access* **10**, 113493–113512. https://doi.org/10.1109/access.2022.3217227 (2022).
45. Qureshi, W. S. et al. Deep learning framework for intelligent pavement condition rating: A direct classification approach for regional and local roads. *Autom. Construct.* **153**, 104945. https://doi.org/10.1016/j.autcon.2023.104945 (2023).
46. Kastryulin, S., Zakirov, J., Prokopenko, D. & Dylov, D. V. Pytorch image quality: Metrics for image quality assessment. https://doi.org/10.48550/ARXIV.2208.14818 (2022).
47. Padilla, R., Netto, S. L. & da Silva, E. A. B. A survey on performance metrics for object-detection algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. 237–242. https://doi.org/10.1109/IWSSIP48289.2020.9145130 (2020).
48. Reis, D., Kupec, J., Hong, J. & Daoudi, A. Real-time flying object detection with yolov8. https://doi.org/10.48550/ARXIV.2305.09972 (2023).

Author contributions

S.C.O and E.S.O conceived and conducted the experiment(s), and analysed the results. S.C.O wrote the manuscript. All authors reviewed the manuscripts.

Funding

This work has been co-financed by the Ministry of Science and Innovation (ES) through the State Plan for Scientific and Technical Research and Innovation 2021-23 under the project MAPSIA [TED2021-129749B-I00]. Also, by the Horizon Europe Research and Innovation Framework program of the European Union under the project LIAISON [101103698].

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.C.-O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.