Received: 17 December 2024 | Accepted: 21 February 2025

DOI: 10.1111/mice.13451

COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING WILEY

RESEARCH ARTICLE

Generative adversarial network based on domain adaptation for crack segmentation in shadow environments

Yingchao Zhang[1] | **Cheng Liu**[1,2]

[1]Department of Systems Engineering, City University of Hong Kong, Hong Kong, China

[2]Centre for Intelligent Multidimensional Data Analysis, City University of Hong Kong, Hong Kong, China

Correspondence
Cheng Liu, Department of Systems Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, China.
Email: cliu647@cityu.edu.hk

Funding information
New Faculty Start-up Fund from the City University of Hong Kong, Grant/Award Number: 9610612; Research Matching Grant Scheme, Grant/Award Number: 9229141; CityU Strategic Interdisciplinary Research Grant, Grant/Award Number: 7020076

Abstract

Precision segmentation of cracks is important in industrial non-destructive testing, but the presence of shadows in the actual environment can interfere with the segmentation results of cracks. To solve this problem, this study proposes a two-stage domain adaptation framework called GAN-DANet for crack segmentation in shadowed environments. In the first stage, CrackGAN uses adversarial learning to merge features from shadow-free and shadowed datasets, creating a new dataset with more domain-invariant features. In the second stage, the CrackSeg network innovatively integrates enhanced Laplacian filtering (ELF) into high-resolution net to enhance crack edges and texture features while filtering out shadow information. In this model, CrackGAN addresses domain shift by generating a new dataset with domain-invariant features, avoiding direct feature alignment between source and target domains. The ELF module in CrackSeg effectively enhances crack features and suppresses shadow interference, improving the segmentation model's robustness in shadowed environments. Experiments show that GAN-DANet improves the crack segmentation accuracy, with the mean intersection over union value increasing from 57.87 to 75.03, which surpasses the performance of existing state-of-the-art domain adaptation algorithms.

1 | INTRODUCTION

Highway networks in various countries around the world are developing faster and provide a good foundation for economic development. However, highways may suffer from cracks or potholes due to rain erosion and vehicle cyclic loads. If these damages are not identified in time, the scope of their influence will be further increased. Cracks may cause rainwater to penetrate into the subgrade structure, leading to material deterioration or reduced structural strength, which can seriously affect traffic safety. Therefore, it is important to detect and repair pavement cracks in time.

Currently, common pavement damage inspection methods include manual inspection, multifunctional defect inspection vehicles and intelligent detection algorithms (Meng et al., 2023; Yingchao Zhang et al., 2021). The manual detection method has high accuracy, but it requires the closure of the detection section, which causes great inconvenience to traffic (Yingchao Zhang et al., 2022). Multifunctional defect inspection vehicles can detect defects in pavements as well as subgrade, but their expensive price

tag prevents them from being utilized in large-scale inspection scenarios. With the development of computer vision, more and more scholars have begun to study the structures' damage detection algorithms based on computer vision (Yingchao Zhang & Liu, 2024c). These models can be used in bridges, dams, airport runways, tunnels, and so on.

Algorithms such as image classification (Hoang & Nguyen, 2019), object detection (F. Yang et al., 2019), and semantic segmentation (J. Liu et al., 2020) are all used to implement intelligent crack inspection systems. The problem of crack image classification has been worked on for many years. Recent research has focused on how to optimize the training process or tune hyperparameters of crack classification networks (Ottoni et al., 2023). Crack detection networks can not only fulfill the task of crack classification but also can detect the exact position of cracks in the image. Therefore, the detection network is more popular at present. A road crack detection model called MUENet was proposed by He et al. (2023), which can solve the problem of similarity between cracks and pavement color well and also achieve better field detection results. G. Zhu et al. (2024) proposed a lightweight encoder–decoder network with hybrid attention and residual blocks for accurate and efficient pavement crack detection, achieving state-of-the-art performance across multiple datasets and real-time detection at 25 FPS with only 0.57 M parameters. Semantic segmentation is a technique to classify the categories of pixels (García-Aguilar et al., 2023; Jianyong Wang et al., 2023). Q. Zhu and Ha (2022) proposed a bidirectional self-rectifying network with Bayesian modeling for improving the segmentation accuracy of cracks. The convolution kernel randomized by Bayesian modeling significantly reduced the sensitivity of the model to uncertainties and nonlinearities, and thus excellent crack segmentation was achieved. In addition, ordinary differential equations (Hu et al., 2023) and dynamic instance generation (Nardin et al., 2023) were also used to improve the performance of segmentation models. Q. Zhu and Ha (2022) proposed a bidirectional self-rectifying network with a Bayesian model to enhance semantic segmentation for surface inspection of built infrastructure, effectively addressing uncertainty and nonlinearity to improve detection accuracy.

Although various models have been developed to detect pavement cracks, environmental factors can have a serious impact on the results when applied in practice (Yingchao Zhang & Liu, 2024b). Yu Zhang and Zhang (2024) used generative adversarial networks (GANs) were used to solve the blurring problem of crack images. This is an environmental factor worth considering. However, shadows caused by buildings or trees on the side of the roadway are the most common influencing factor. Crack detection in shaded environments is challenging because it is difficult for the model to distinguish the edges of cracks from shadows. Shadows can significantly reduce the pixel contrast between the crack and the background. In low-contrast situations, models may not be able to identify the edges and details of cracks efficiently. Therefore, it is challenging to segment crack pixels with high accuracy in shadow environments. L. Fan et al. (2023) proposed a two-stage method for crack segmentation in shaded environments. This method primarily mitigates the impact of shadow noise on crack segmentation by employing shadow removal and data augmentation techniques that leverage differences in luminance values. Yingchao Zhang and Liu (2024a) developed a one-stage approach to crack segmentation in shaded environments by integrating the process of shadow removal into the network. A detection model called CrackcellNet was proposed by H. Yang et al. (2024), which contains a module to enhance shadow image data, improving the F1 value for crack detection in shadowed environments to 0.839. In addition, neural dynamic classification algorithm (Rafiei & Adeli, 2017), finite element machine (Pereira et al., 2020), dynamic ensemble learning algorithm (Alam et al., 2020), self-supervised learning (Rafiei et al., 2022, 2024) all have the potential in the field of roadway defect segmentation, but their application in crack segmentation is currently less common and therefore will not be continued.

The aforementioned scholars are typically engaged in the design of novel network architectures or modules with the objective of addressing the challenge of crack segmentation in shaded environments. They have also employed fully supervised learning techniques to achieve highly satisfactory outcomes. However, the acquisition of a substantial number of shaded datasets represents a significant challenge. The number of samples in the publicly available crack datasets is approximately several hundred to more than a thousand, of which only a few dozen contain shadows. To obtain a better segmentation effect in the shadowed environment, 800 to 1000 crack images in the shadowed environment are generally required. Therefore, GAN is commonly used to enhance data (Fu et al., 2024; Leng et al., 2024). Zhong et al. (2024) used the residual attention mechanism, a multi-field perceptual discriminator, and a self-encoder perceptual loss to improve the performance of GAN. Although H. Zhang et al. (2024) proposed a crack diffusion model to generate the dataset of cracks under shadows or water traces, this dataset differs significantly from real cracks. Because the sharp edges and uniform patterns of the generated shadows differ from natural shadows, this may limit the realism of the synthetic dataset. And natural shadows exhibit irregular shapes, soft edges, varying intensity, and dynamic interactions with surface textures. These characteristics

make them challenging to replicate in synthetic data, yet critical for improving the generalization of segmentation models in real-world environments. Xue et al. (2023) combined evolutionary algorithms and attention mechanisms to improve CycleGAN training stability and image generation quality by optimizing generator configurations and enhancing feature learning.

Some scholars also used semi-supervised learning to solve the problem of insufficient data. Díaz-Francés et al. (2024) proposed a novel deep learning architecture that combines ideas from EditGAN and generative diffusion models to enhance semi-supervised image segmentation, achieving improved performance on multiclass and binary segmentation tasks.

In summary, current research employs three main types of methods to achieve crack detection in shaded environments:

1. Two-stage shadow removal approach: This method first preprocesses images to eliminate shadows before conducting crack detection. While straightforward to implement, this approach risks losing crucial crack detail information during the shadow removal process.
2. End-to-end single-stage approach: This strategy incorporates shadow removal directly within the detection network architecture. Although this method preserves feature integrity, it requires extensive training data containing shadowed examples, which can be challenging to obtain.
3. Data augmentation-based methods: These techniques expand training datasets by artificially generating shadowed images. However, current shadow generation methods often produce unrealistic results, with edge characteristics and distribution patterns that deviate significantly from natural shadows.

In addition, these methods mainly focus on global feature alignment, ignoring the importance of local edge information in the crack detection task.

Therefore, many scholars began to explore the potential of transfer learning in the domain of crack segmentation. Transfer learning based on shared parameters has more applications in the field of crack detection or segmentation (Q. Yang et al., 2020). Dais et al. (2021) improved the segmentation accuracy of the cracks by 6% by fine-tuning the pre-trained weights. Katsigiannis et al. (2023) used pre-trained weights obtained from a large dataset to classify wall cracks. Transfer learning enabled the accuracy to reach 99%. However, it is more difficult to obtain pre-trained weights in large datasets such as common objects in context dataset (COCO) or visual object classes (VOC) when a novel model is proposed. To solve this problem, Çelik and König (2022) proposed a novel transfer learning method. By using weights obtained from training on a dataset containing 1000 images as pre-trained weights, the accuracy on different datasets can be significantly improved. Transfer learning based on shared parameters is relatively simple and easy to implement, especially when the source and target domains are quite similar. However, this approach also has some limitations. The parameters of the source domain model may contain too many features specific to the source domain rather than general features of cracks. This can lead to overfitting in the target domain during transfer learning.

Domain adaptation-based transfer learning is specifically used to handle the distribution differences between the source and target domains. Among these methods, GAN (Goodfellow et al., 2020) is a common approach. GANs are commonly used by scholars to generate new datasets or increase sample sizes (T. Zhang et al., 2023). Maeda et al. (2021) proposed using GANs and Poisson blending to generate realistic synthetic road damage images, enhancing training data and improving road damage detection accuracy. Gao et al. (2021) introduced the balanced semi-supervised GAN, a GAN variant designed to address data deficiency and class imbalance in structural health monitoring, achieving superior performance in concrete damage detection through balanced-batch sampling and semi-supervised learning. The method for synthesizing crack images using a conditional GAN and a self-training approach to enhance crack detection was proposed by Shim (2024), addressing data scarcity and achieving improved performance with an mean intersection over union (mIoU) of 80.34% and F1-score of 76.31%. A transfer learning pipeline was proposed in Yishun Li et al. (2021), where data transfer with GAN-based scene style adaptation and model transfer via domain adaptation are combined, enabling pavement distress detection models to be applied across new scenarios, with training data needs reduced by 25% and accuracy improved by 26.55%. A self-training framework based on a Bayesian neural network and spatial prior was proposed to realize high-precision automatic segmentation of concrete wall cracks by unsupervised domain adaptation, which can significantly improve the model performance with only 100 images of the target domain (Chun & Kikuta, 2024). X. Fan et al. (2022) designed a GAN to enable the crack segmentation model to learn features from both the source domain and the target domain simultaneously. When the GAN cannot distinguish whether the data come from the source domain or the target domain, the crack segmentation task based on transfer learning is achieved. Ma et al. (2023) developed a shape-consistent style transfer module that performs pixel-level distributions between training and test images so that the pixel distributions of the training and test sets are consistent. This domain-adaptive

approach substantially enhances the model's robustness against variations in crack distributions and achieves superior crack segmentation.

In real-world inspection scenarios, existing computer vision-based inspection methods encounter significant limitations. A critical challenge arises from shadows cast by buildings and trees onto pavements, which severely impair crack detection accuracy. These shadows substantially reduce the pixel contrast between cracks and their surrounding background. Moreover, the textural characteristics of shadowed regions closely resemble those of actual cracks, frequently resulting in false positive detections. Based on the preceding analysis, while existing methods employing GAN to generate shadow images can augment datasets, the generated shadows often lack natural realism and fail to accurately replicate shadow characteristics observed in real-world scenarios. Furthermore, during shadow generation, the edge and detail information of cracks becomes blurred, hindering the model's ability to effectively capture the integrity of crack features during training. And domain-adaptive approaches primarily focus on aligning the global feature distributions between the source and target domains, often neglecting the fine-grained features critical to the crack segmentation task. This global alignment results in suboptimal extraction of local boundary information for cracks in shaded environments. Given the critical importance of accurate crack detection in shadowed environments and the limitations of current research approaches, there is a pressing need to develop robust detection methods capable of operating effectively under varying lighting conditions. GANs have demonstrated remarkable capabilities in image synthesis and domain adaptation tasks. Furthermore, domain adaptation techniques offer promising solutions to address domain shift challenges, making them particularly relevant for this application. Therefore, this study proposed an innovative method that synergizes GAN-based shadow image generation with domain adaptation, aiming to leverage the advantages of both approaches while mitigating their respective shortcomings. This method achieves better detection by generating a new dataset with domain invariance to overcome the problem of direct feature alignment while maintaining the local detailed features of the cracks. The main contributions of this paper can be summarized as follows:

1. This paper proposed a GAN for fusing unshaded and shaded datasets to reduce the distance between the target and source domain. The crack segmentation network was trained on the newly generated dataset and excellent segmentation results can be obtained.
2. The crack segmentation network incorporated enhanced Laplacian filtering (ELF). Most of the shadow features were filtered out during feature extraction, thus achieving excellent transfer effects in the target domain.
3. The method was validated using actual datasets with limited shadowed images, confirming its robustness and applicability in data-scarce environments.

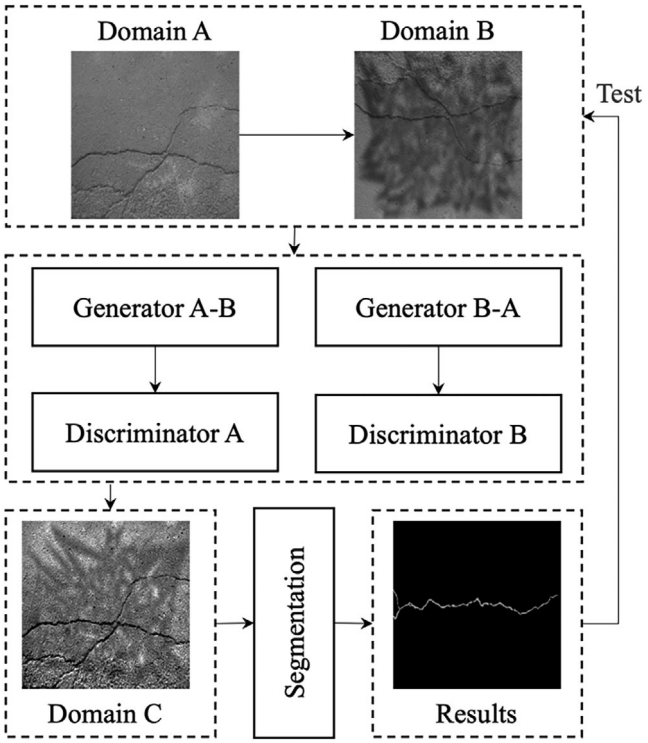

FIGURE 1 Structures of GAN-DANet networks for pavement crack segmentation in shaded environments.

The rest of this paper is organized as follows: Section 2 focuses on the research methodology proposed in this paper. Section 3 introduces the crack datasets. Sections 4 and 5 present the Results and Discussion and the Conclusion, respectively.

2 | METHODOLOGY

2.1 | Segmentation framework

To effectively detect pavement cracks in shaded environments, this research proposed a novel two-stage segmentation algorithm named generative adversarial networks and domain adaption network (GAN-DANet), which can be seen in Figure 1. The first stage of GAN-DANet primarily consists of CrackGAN. Due to the challenge of acquiring a large number of images of cracks in shaded environments, this study used 4776 real crack images and 100 artificial shadowed images as inputs to CrackGAN,

TABLE 1 Definition of model components.

Symbol	Definition
Domain A (Source domain)	Datasets without shadows
Domain B (Target domain)	Datasets with shadows
$G_{A\to B}$	Convert the image of domain A to domain B
$G_{B\to A}$	Convert the image of domain B to domain A
D_A	Distinguish between the real image of domain A and the generated image
D_B	Distinguish between the real image of domain B and the generated image

which generated a new dataset of realistic shadowed crack images. These generated images were used as the training set for the segmentation network. A comprehensive description of the dataset is provided in Section 3. In Figure 1, the process begins with two distinct datasets: domain A, which contains crack images captured under normal lighting conditions without shadows (top-left), and domain B, which includes crack images heavily influenced by complex shadow patterns (top-right). These datasets are utilized as input to the CrackGAN to produce a synthesized domain C. Domain C is designed to combine the characteristics of both datasets, effectively bridging the gap between shadow-free and shadowed environments. The generated domain C is then used to train the proposed semantic segmentation model, which is optimized to perform crack segmentation with high accuracy, even under challenging shadow conditions. The bottom-right image demonstrates the binary crack map produced by the segmentation model. In addition, to evaluate the robustness of GAN-DANet, the trained segmentation model is directly tested on the original datasets (domains A and B). Despite the significant differences in lighting conditions between the two datasets, our model achieves excellent performance on both. The detailed structure of CrackGAN and CrackSeg is described in the next section. The methods used in this study were chosen based on their effectiveness in addressing shadow interference in crack segmentation, computational efficiency, and practical applicability to domain-specific datasets.

2.2 | Structures of CrackGAN

As shown in Figure 2, CrackGAN mainly consists of two generators ($G_{A\to B}$ and $G_{B\to A}$) and two discriminators (D_A and D_B). Before the introduction of the model, this research defined the symbols as shown in Table 1. The generator is mainly used to realize the migration between the source and target domains, while the discriminator is used to determine whether the input image is from a real image or a generated image. When the images produced by the generator are highly realistic and the discriminator cannot distinguish them well, it suggests that the CrackGAN model has reached a state of convergence. The components of CrackGAN are detailed sequentially in the sections below.

2.2.1 | Generator

The input images were initially normalized to adjust the pixel values to a range between −1 and 1 to get the input feature maps for the generator. Subsequently, these feature maps were processed through the CIR@7 module, which utilizes 7×7 large convolutional kernels to capture more contextual information. In addition, instance normalization layer was also used to improve the stability and convergence speed of training. The output feature maps were sequentially processed by two CIR@3 modules with nine base block modules. Convolutional kernels of 3 × 3 were used to extract features incrementally and reduce the computational cost of the network in the CIR@3 architecture. Base block used residual structures based on CIR@3 to alleviate the problem of vanishing gradients of deep neural networks. The CTIR@3 module utilized transposed convolution for upsampling, effectively increasing the resolution of the feature maps. This is primarily because transposed convolution has learnable parameters to learn complex spatial feature transformations for detailed feature reconstruction. Finally, feature maps were processed by the convolutional layer of 7×7 convolutional kernels with an activation layer to get the generated images. The generator of CrackGAN adopts a global-to-local feature extraction strategy, where CIR@7 is used first to capture broad contextual information with a larger kernel size, followed by CIR@3 to refine local details with a smaller kernel size. CIR@7 is critical for distinguishing cracks from shadows and other large-scale patterns, while CIR@3 enhances the resolution of fine-grained features, such as thin or faint cracks. CTIR@3 performs feature reconstruction and upsampling, increasing the spatial resolution of the feature maps and recovering fine details. This hierarchical design ensures that the generator can handle cracks of varying scales and appearances effectively.

The input and output of the generator $G_{A\to B}$ are the shadow-free crack images and generated images with

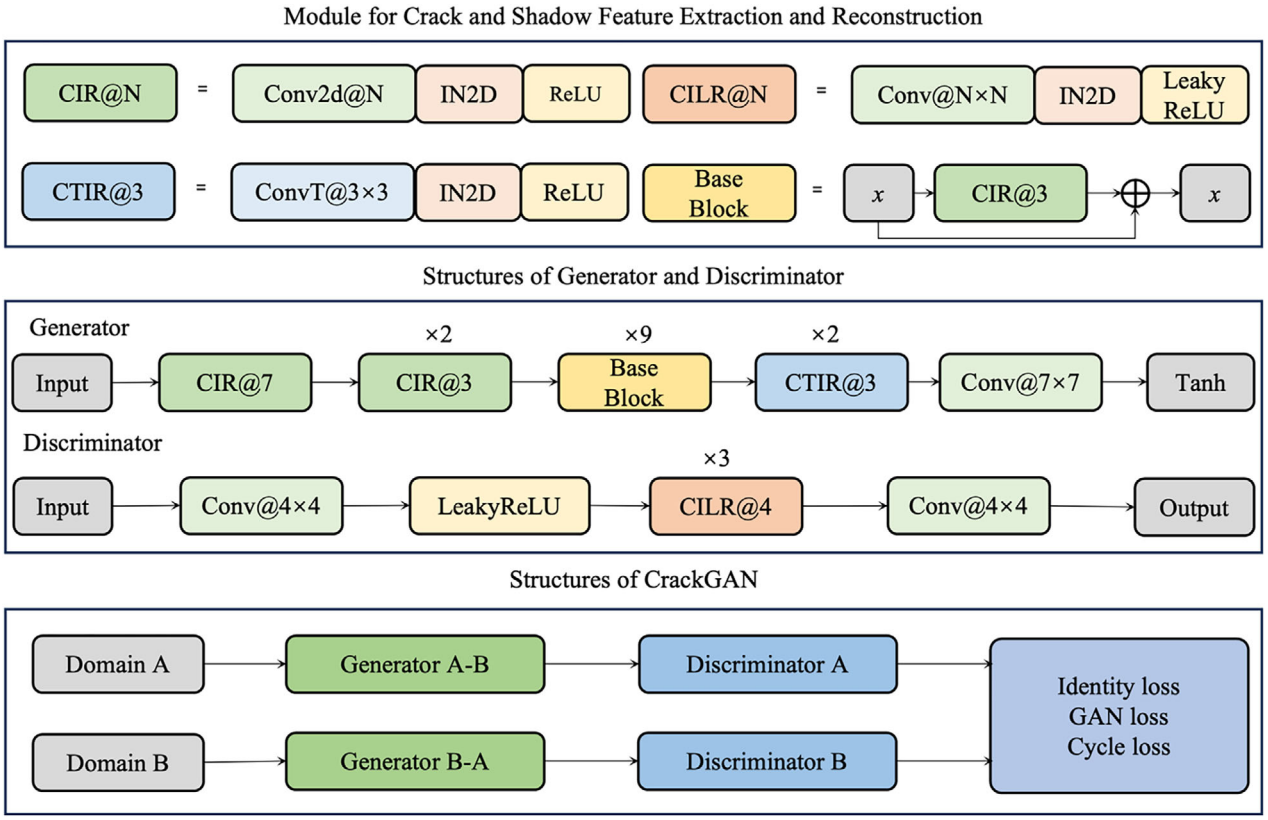

FIGURE 2 Structures of CrackGAN. The picture shows the generator and discriminator structure of CrackGAN. The generator comprised an input layer, multiple CIR@N and CTIR@3 layers, nine base blocks, and a tanh output layer. The discriminator consisted of an input layer, a Conv@4×4 layer, an activation layer, CILR@4 layers, and an output layer.

fused shadow features, respectively. The input of generator $G_{B\to A}$ is real shadow images, and the output is generated images with fewer shadow features. A total of three loss functions are used in the image generation phase to ensure that CrackGAN can be efficiently transformed between different domains, including Identity loss, GAN loss, and cycle loss.

Identity loss is mainly used to keep the color or other basic features consistent. In particular, when an image of the target domain is fed into the generator, the generator should be able to output an image that is similar to the input. The formulas are shown in Equations (1) and (2) (J.-Y. Zhu et al., 2017), where x and y denote the data in domains A and B, respectively. This loss allows the model to maintain the parts of the image that do not need to be changed when performing a style transformation.

$$\mathcal{L}_{id}^{A} = E_{x\sim p_{data}(x)}\left[\| G_{B\to A}(x) - x\|_1\right] \quad (1)$$

$$\mathcal{L}_{id}^{B} = E_{y\sim p_{data}(y)}\left[\| G_{A\to B}(y) - y\|_1\right] \quad (2)$$

where $\mathcal{L}_{id}^{A}$ and $\mathcal{L}_{id}^{B}$ denote the identity loss, which measures the ability of the generator to maintain the original features of the image. $G_{B\to A}(x)$ represents the generator that transforms from domain B to domain A, and $G_{A\to B}(y)$ is the generator that transforms from domain A to domain B. x and y are the input source and target domain images, respectively. E represents the expected value.

GAN loss is mainly used to help the generator to improve the quality of the generated images. The formulas are shown in Equations (3) and (4) (J.-Y. Zhu et al., 2017), where $D_B(G_{A\to B}(x))$ and $D_A(G_{B\to A}(x))$ represent the output of the discriminator to the generated images. The value of $D_B(G_{A\to B}(x))$ or $D_A(G_{B\to A}(x))$ is close to 1 when the images generated by the generator are considered by the discriminator to be real images. This allows the generator to produce images that can trick the discriminator. By minimizing this loss, the generator is able to improve the quality of the generated images.

$$\mathcal{L}_{GAN}^{A\to B} = E_{x\sim p_{data}(x)}\left[\left(D_B\left(G_{A\to B}(x)\right) - 1\right)^2\right] \quad (3)$$

$$\mathcal{L}_{GAN}^{B\to A} = E_{y\sim p_{data}(y)}\left[\left(D_A\left(G_{B\to A}(y)\right) - 1\right)^2\right] \quad (4)$$

where $\mathcal{L}_{GAN}^{A\to B}$ and $\mathcal{L}_{GAN}^{B\to A}$ are the adversarial loss function. $A\to B$ represents a transition from an unshaded domain to a shaded domain, while $B\to A$ does the opposite. D_A and D_B

are discriminators used to determine whether the image is real or generated.

Cycle loss is mainly used to assist the model to learn the ability to convert between different domains without losing important information. Specifically, when images were transformed from domain A to domain B and then from domain B back to domain A, the final images obtained should be as similar as possible to the original images. The formulas are shown in Equations (5) and (6) (J.-Y. Zhu et al., 2017).

$$\mathcal{L}_{cyc}^{A\to B\to A} = E_{x\sim p_{data}(x)}\left[\| G_{B\to A}\left(G_{A\to B}\left(x\right)\right) - x\|_1\right] \quad (5)$$

$$\mathcal{L}_{cyc}^{B\to A\to B} = E_{y\sim p_{data}(y)}\left[\| G_{A\to B}\left(G_{B\to A}\left(y\right)\right) - y\|_1\right] \quad (6)$$

where $\mathcal{L}_{cyc}^{A\to B\to A}$ and $\mathcal{L}_{cyc}^{B\to A\to B}$ are the cycle loss.

The total loss in the generator is shown in Equation (7) (J.-Y. Zhu et al., 2017). To adjust the different loss terms to the same magnitude to avoid one loss term dominating the training, this research set the weighting coefficients λ_{id} and λ_{cyc}. The parameters λ_{id}, and λ_{cyc} were manually tuned within the ranges 0–1 and 1–20, respectively. The results were not highly sensitive to these parameters. Performance metrics varied only slightly (e.g., within ±0.5% for mIoU and ±0.5% for mean pixel accuracy [mPA]) when these parameters were adjusted within their respective ranges. Therefore, λ_{id} equals to 0.5, and λ_{cyc} equals to 10.

$$\begin{aligned}\mathcal{L}_T = {} & \mathcal{L}_{GAN}^{A\to B} + \mathcal{L}_{GAN}^{B\to A} + \lambda_{id}\left(\mathcal{L}_{id}^{A} + \mathcal{L}_{id}^{B}\right) \\ & + \lambda_{\text{cyc}}\left(\mathcal{L}_{cyc}^{A\to B\to A} + \mathcal{L}_{cyc}^{B\to A\to B}\right)\end{aligned} \quad (7)$$

where $\mathcal{L}_T$ *is* the total loss.

2.2.2 | Discriminator

The structure of the discriminator is simpler, compared to the generator, because it only needs to determine whether the image is real or generated. First, input images went through a 4×4 convolutional layer for initial feature extraction and were processed by a leaky rectified linear unit (LeakyReLU) activation function layer. The output feature maps were then processed by three CILR@4 modules. Each CILR@4 module contains a 4×4 convolutional layer, instance normalization layer, and activation function layer. The sequential stacking of the three modules progressively captures the multilevel features of the images. Finally, the output of the discriminator is the probability value that the input image is a real image. The losses during the training of the two discriminators are shown in Equations (8) to (13) (J.-Y. Zhu et al., 2017).

$$\mathcal{L}_{\text{real\_A}} = E_{x\sim p_{data}(x)}\left[\left(D_A\left(x\right) - 1\right)^2\right] \quad (8)$$

$$\mathcal{L}_{\text{fake\_A}} = E_{x\sim p_{data}(x)}\left[\left(D_A\left(G_{B\to A}\left(y\right)\right)\right)^2\right] \quad (9)$$

$$\mathcal{L}_{D_{\text{A}}} = \left(\mathcal{L}_{\text{real\_A}} + \mathcal{L}_{\text{fake\_A}}\right)/2 \quad (10)$$

$$\mathcal{L}_{\text{real\_B}} = E_{y\sim p_{data}(y)}\left[\left(D_B(y) - 1\right)^2\right] \quad (11)$$

$$\mathcal{L}_{\text{fake\_B}} = E_{y\sim p_{data}(y)}\left[\left(D_B\left(G_{A\to B}\left(x\right)\right)\right)^2\right] \quad (12)$$

$$\mathcal{L}_{D_{\text{B}}} = \left(\mathcal{L}_{\text{real\_B}} + \mathcal{L}_{\text{fake\_B}}\right)/2 \quad (13)$$

where $D_A(x)$ represents discriminator A's output for real source domain image x, and $D_B(x)$ represents discriminator B's output for real target domain image y. $\mathcal{L}_{D_A}$ and $\mathcal{L}_{D_B}$ are the total loss for discriminators A and B, respectively.

2.3 | Structures of CrackSeg

Given the high accuracy of current semantic segmentation networks, this study introduces the CrackSeg model for shadow segmentation, which is based on high-resolution net (HRNet; Jindong Wang et al., 2020). The structure of CrackSeg is shown in Figure 3.

HRNet is used as the segmentation model for cracks because of its powerful feature representation. HRNet maintains high-resolution feature maps during feature extraction while introducing low-resolution feature maps. The feature maps of varying resolutions are continuously fused, thereby enabling the model to capture richer contextual information while retaining texture information. This is extremely important for crack segmentation, which is sensitive to edge and texture information. In addition, ELF (F. Zhang et al., 2024) was incorporated into the network to augment the edge and detail information in the image, which can be seen in Equations (14) to (15). In crack edge feature detection, the second-order derivative of pixels is computed using the Laplace operator to emphasize areas with rapid changes, effectively filtering out regions where the pixel value changes are not obvious. In image processing, low-frequency information includes large uniform color blocks and gradually changing color gradients, whereas high-frequency information comprises primarily edge and texture information. The ELF models were used in CrackSeg mainly to filter out low-frequency information and retain high-frequency information. The ELF models were embedded in the backbone of the model in order to filter out redundant features as early as possible.

$$\Delta I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (14)$$

$$I_{enhanced} = I + \alpha \cdot \Delta I \quad (15)$$

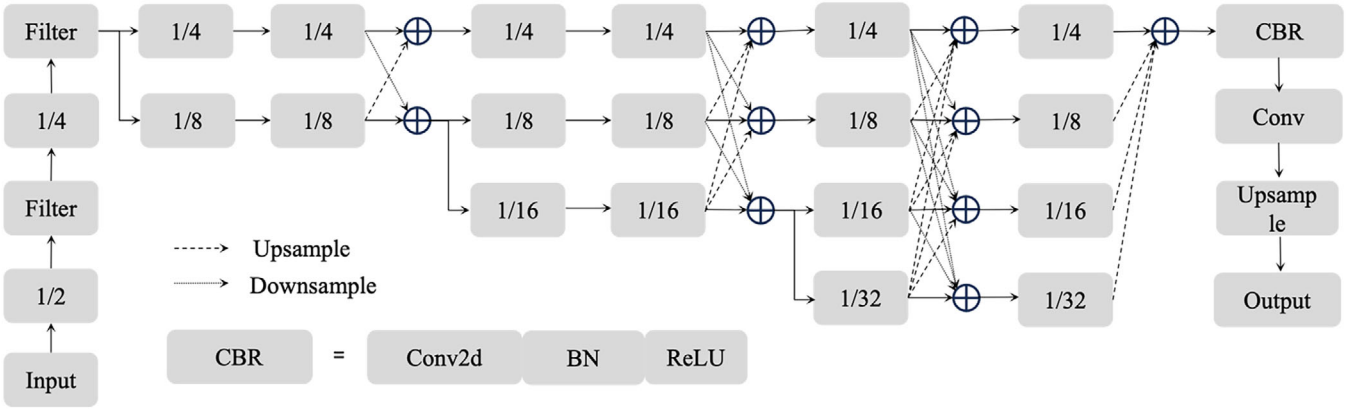

FIGURE 3 Structures of CrackSeg. $1/n$ means that the width and height of the feature maps are $1/n$ of the input image.

where I denotes the input feature maps, ΔI is the result of Laplace operator processing, and α indicates the enhancement coefficient. The parameter α was manually tuned within the range 1–5. In this study, the model has better performance when α is set to 2. Various experimental results have demonstrated that it achieves the best outcomes.

3 | DATASETS

The main purpose of this study is to improve cracks segmentation accuracy under shaded environments. In the following, the dataset used by this paper will be introduced.

3.1 | Crack dataset without shadows

The primary datasets commonly utilized for the semantic segmentation of cracks include Crack500 (F. Yang et al., 2019), CrackForest dataset (Shi et al., 2016), and DeepCrack (Y. Liu et al., 2019). Various horizontal cracks, longitudinal cracks, and alligator cracks are included. Since the focus of this research is on improving crack segmentation accuracy in shaded environments through domain adaptation rather than transfer learning across different datasets without shadows, this research combined three datasets to obtain a large dataset for cracks segmentation without shadows.

In total, this research obtained 4776 images containing pavement cracks. As these images originate from three different datasets with varying resolutions, this research adjusted the resolution of each image to 512×512 pixels to ensure consistency. The dataset obtained after being adjusted is represented in this study by D1 as shown in Figure 4.

3.2 | Crack dataset with shadows

Obtaining a large, high-quality dataset of real-world shadowed road distress images is a challenging and resource-intensive task. Due to the lack of publicly available datasets under shadowed conditions, this study adopted a synthetic dataset generation approach to simulate shadow patterns. This approach enables us to train the model effectively while addressing the scarcity of annotated data under such conditions. Therefore, this study utilizes the generated cracks dataset with shadows from Yingchao Zhang and Liu (2024a). These images were created by mixing the generated shadow layers with the standard crack images. Compared to traditional shadow generation methods, which often rely on fixed patterns or geometric models, CrackGAN generates shadows that are more diverse and realistic, featuring irregular shapes, varying intensities, and soft transitions. Additionally, it preserves the structural integrity of cracks, ensuring that the generated images remain suitable for crack segmentation tasks.

The process of generating shadows includes the following steps (Yingchao Zhang & Liu, 2024a): Initially, a transparent layer matching the original image's dimensions is generated, followed by defining the shadow's color and transparency. The transparency of the shadows in this study was set to 30. Next, ellipse, rectangle, and polygon shapes are randomly generated and drawn on the transparent layer. Finally, a Gaussian blur is applied to the shadow layer to soften the edges of the shadows, making them more closely resemble actual shadows. This research called this dataset D2 as shown in Figure 5. A visual inspection of the artificial shadows reveals that they closely resemble the overall appearance of real-world shadows. For example, elliptical shadows emulate those cast by roadside traffic signs, while polygonal shadows mimic the irregular patterns of shadows caused by tree branches. These design

FIGURE 4 Instances of D1. Examples of pavement crack datasets without shadows.

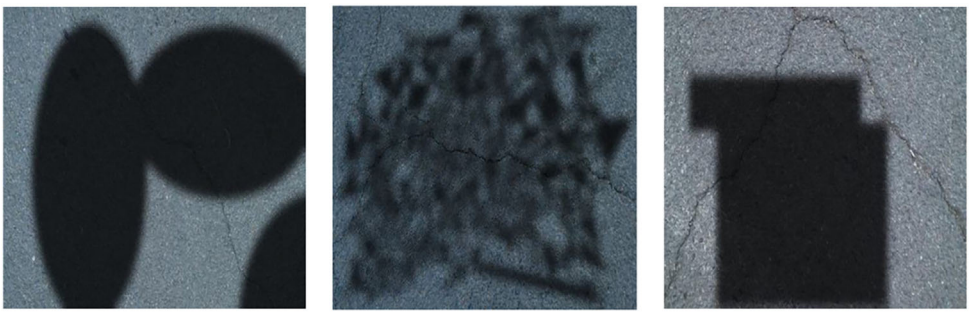

FIGURE 5 Instances of D2. Examples of pavement crack datasets with shadows.

choices ensure that the artificial shadows introduce realistic shadow-induced challenges into the dataset, such as reduced pixel contrast and obscured crack edges.

This research compared the pixel frequency distributions of natural shadows, artificial shadows, and images without shadows as shown in Figure 6. The histogram is used to compare pixel intensity distributions among different images. This allows this study to evaluate the fidelity of the generated images in terms of pixel-level intensity preservation and identify any potential artifacts or deviations. The image without shadows exhibits more uniform brightness, characterized by a sharp peak, indicating the absence of distinctly dark or bright pixels. Images with natural shadows display a peak in regions of lower brightness, suggesting an increased presence of dark pixels or shadows. For images with artificially generated shadows, while they lack the clear bimodal distribution characteristic of natural shadows, these artificial shadows still influence the image's gray value distribution, shifting it toward lower gray values. The gray value distributions of both natural and artificial shadows are significantly different from those of the unshaded image, exhibiting the expected shadow characteristics. In addition, while artificial shadows may lack the full complexity of dynamic real-world shadows, their design effectively captures the essential characteristics needed to train domain adaptation models. Therefore, this research can consider these generated shadows to be valid. In this process, the total number of images of cracks in shadow generated was 100.

In total, this research obtained 4776 images containing pavement cracks. As these images originate from three different datasets with varying resolutions, this research adjusted the resolution of each image to 512×512 pixels to ensure consistency. The dataset obtained after being adjusted is represented in this study by D1 as shown in Figure 4.

4 | RESULTS AND DISCUSSION

4.1 | Training environment and hyperparameters

The various software and hardware versions used during training are shown in Table 2. The stochastic gradient descent optimizer was used to update the parameters of the model during the training process. The initial learning rate was set at 0.004 and was subject to cosine decay, with a minimum value established at 0.00004. The cosine decay algorithm smoothly adjusts the learning rate to avoid

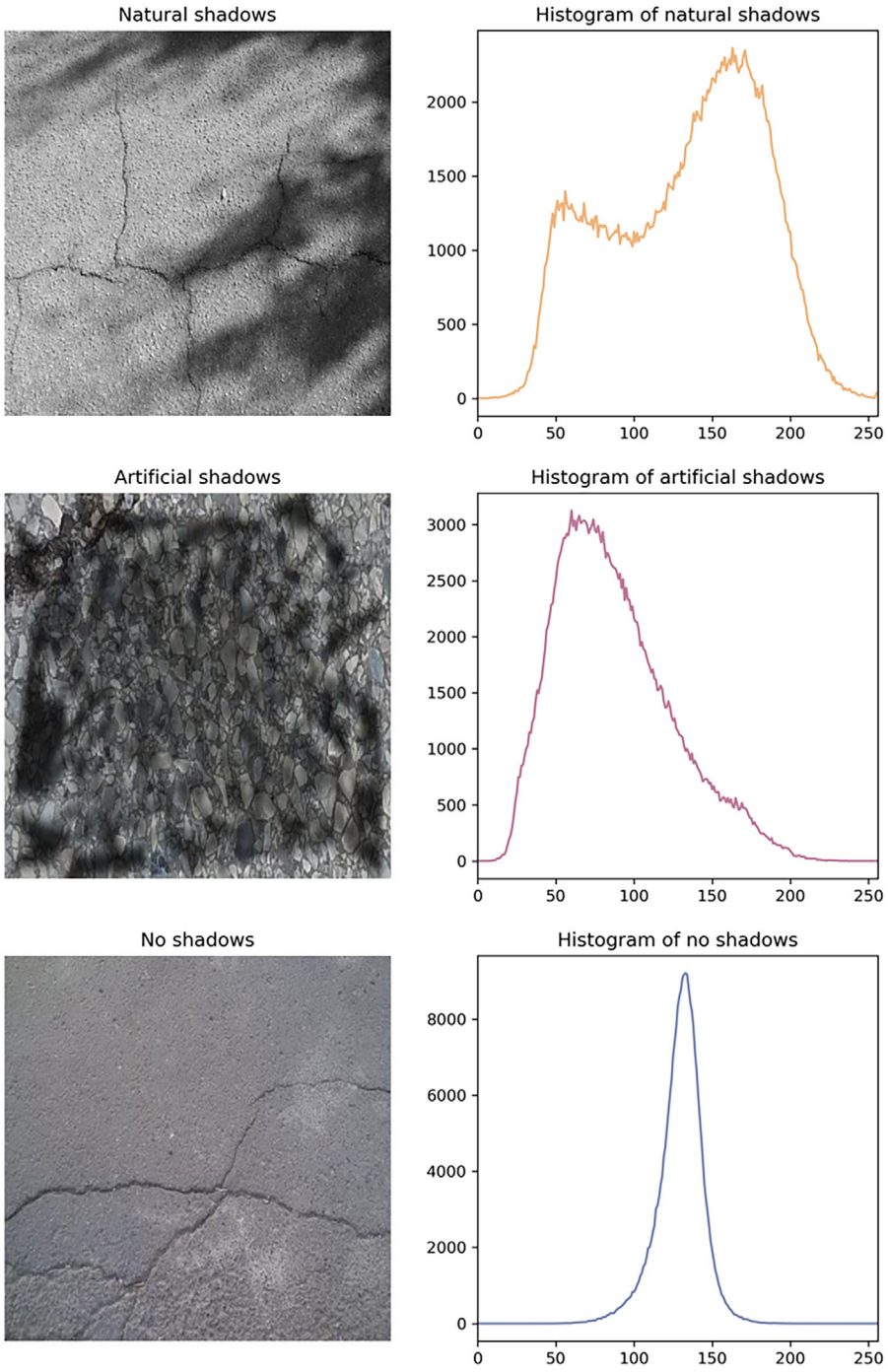

FIGURE 6 Histogram of pixel frequency distribution for different images.

oscillations in the model during training. In summary, the models were all trained to convergence according to the hyperparameter settings in Table 2.

When training the CrackGAN, only 100 pavement damage images in shaded environments were used as the training set. This is significantly fewer than what is typically used in fully supervised learning methods for pavement damage detection in similar conditions (Yingchao Zhang & Liu, 2024a).

In addition, this study addresses a domain adaptation problem. The model is trained exclusively on domain A and evaluated on both domain A and domain B. Domain B represents an unseen domain with a different data distribution, which naturally results in lower performance due to domain shift. To ensure the reliability of the reported results, this study averaged the metrics over five independent runs with different random seeds, accounting for fluctuations during the training process.

TABLE 2 The versions of software and hardware.

Item	Version	Item	Version
CPU	Intel i9-13900K	Optimizer	SGD
GPU	2 × NVIDIA RTX 4090	Learning rate	4e-3
OS	Ubuntu 20.04	Min learning rate	4e-5
RAM	128 GB	Learning rate decay	Cosine decay
PyTorch	2.2.1	Epoch	200
Python	3.8.18	Momentum	0.9
CUDA	12.1	Batch size	8

Abbreviation: CUDA, Compute unified device architecture. SGD, stochastic gradient descent.

4.2 | Evaluation metrics

This study addresses the semantic segmentation of cracks, so mIoU, mPA, and accuracy were selected as the evaluation criteria to evaluate the model's performance. The mIoU is mainly used to measure the degree of overlap between the predicted and the true region and is calculated as shown in Equation (16). mPA represents the average ratio of each pixel being correctly classified. Compared to mIoU, mPA more directly reflects the pixel-level accuracy of the model. The formula is shown in Equation (17). The accuracy indicates the proportion of all correctly classified pixels to the total pixels. The formula is shown in Equation (18). Although accuracy provides a performance metric that can be calculated quickly, it may not be sensitive enough in the case of class imbalance. This is because even if a few categories are completely misclassified, as long as the majority of categories are classified correctly, the accuracy may still be high, which is particularly true in the field of crack detection in shaded environments. Therefore, the combination of the three metrics allows for a more comprehensive assessment of the model's performance.

$$mIoU = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|Y_i \cap \hat{Y}_i\right|}{\left|Y_i \cup \hat{Y}_i\right|} \tag{16}$$

$$mPA = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|Y_i \cap \hat{Y}_i\right|}{|Y_i|} \tag{17}$$

$$Accuracy = \frac{\sum_{i=1}^{N}|Y_i \cap \hat{Y}_i|}{\sum_{i=1}^{N}|Y_i|} \tag{18}$$

where N represents the number of categories, Y_i denotes the set of true pixels belonging to category i, and $\hat{Y}_i$ is the set of pixels predicted to belong to category i.

4.3 | Experimental results and analysis

4.3.1 | Comparison with baseline

The segmentation results of the proposed GAN-DANet are illustrated in Table 3. HRNet-A denotes the model trained on domain A, HRNet-B signifies training on domain B, and GAN-DANet is trained on domain C, which consists of data generated by CrackGAN. The models that had reached a state of convergence were evaluated using the test sets from domains A and B. The performance of HRNet-A highlights the substantial influence of shadows on the accuracy of crack segmentation. When domain B served as the test set, the mIoU for the HRNet-A model declined from 77.51% to 57.87%, which indicated that HRNet-A could hardly distinguish the cracked pixels from the shadows. The best segmentation results were achieved when the model was trained in domain B, but the values of mIoU as well as mPA dropped significantly when tested directly in domain A. This is because HRNet-B performed fully supervised learning in domain B, and it can achieve high accuracy in the shaded test set. Due to the differing distributions between domain B and domain A, the test results in domain A are inferior. Additionally, domain A does not suffer from shadow interference, resulting in less degradation, compared to HRNet-A. When domain C was utilized as the training set, the model demonstrated robust performance and stability in both domain A and domain B. The results show that GAN-DANet can obtain comparable fully supervised learning model accuracy with a small number of shaded images. It further proves the effectiveness of the algorithm in this study.

GAN-DANet demonstrates robust performance in both shadow-free (domain A) and shadowed (domain B) environments due to its ability to effectively extract crack features under varying conditions. In shadow-free images, GAN-DANet enhances fine-grained details, such as small or faint cracks, through the ELF module, which emphasizes high-frequency edge information while suppressing redundant low-frequency background

TABLE 3 Segmentation results of different models.

Model	Domain A			Domain B		
	Mean intersection over union (mIoU)	Mean pixel accuracy (mPA)	Accuracy	mIoU	mPA	Accuracy
HRNet-A	77.51	85.08	97.71	57.87	66.86	94.21
HRNet-B	72.61	77.39	97.03	75.38	81.12	97.00
GAN-DANet	75.06	80.77	97.23	75.03	80.74	96.95

Note: HRNet-A refers to training in domain A, HRNet-B denotes training in domain B, and GAN-DANet represents training in domain C generated by CrackGAN and testing in domain A and domain B.
Abbreviation: HRNet, high-resolution net.

TABLE 4 Comparison with domain adaptation algorithms.

Model	Domain A			Domain B		
	mIoU	mPA	Accuracy	mIoU	mPA	Accuracy
HRNet-A	77.51	85.08	97.71	57.87	66.86	94.21
MMD	76.13	81.87	97.41	63.92	68.06	94.11
WMMD (Yan et al., 2017)	76.24	82.69	97.46	65.63	71.41	95.83
BDL (Yunsheng Li et al., 2019)	75.71	81.2	97.31	68.11	73.21	96.33
ADPL (Cheng et al., 2023)	76.03	81.52	97.39	72.72	77.65	96.71
GAN	74.86	80.96	97.17	65.87	70.01	96.01
GAN-DANet	75.06	80.77	97.23	**75.03**	**80.74**	**96.95**

Note: Bold values refer that proposed model is superior in these metrics.
Abbreviations: APDL, adaptive dual path learning; GAN, generative adversarial network; HRNet, high-resolution net; MMD, maximum mean discrepancy; WMMD, weighted maximum mean discrepancy.

features. In shadowed images, the data generated by CrackGAN enable the model to learn domain-invariant features, effectively distinguishing cracks from shadows. The ELF module further enhances this capability by accurately preserving crack edges while filtering out shadow interference. This combination of data augmentation and high-frequency feature enhancement gives GAN-DANet a significant advantage over models trained solely on single-domain data, such as HRNet-A or HRNet-B.

4.3.2 | Comparison with state of the art

In this section, GAN-DANet was used to compare with several classical domain adaptation algorithms, and the results are shown in Table 4. To provide a fair comparison, this research employed a series of algorithms that are commonly utilized in the domain adaptation field, including maximum mean discrepancy (MMD), weighted MMD (WMMD), bidirectional learning (BDL) (Yunsheng Li et al., 2019), adaptive dual path learning (Cheng et al., 2023), and GAN. MMD and WMMD mainly work by comparing the output features of HRNet with those of the source and target domains and calculating the differences between them. Their losses were subsequently incorporated into the total loss, guiding the model to learn features that are more invariant across domains. BDL achieved the state of the art in semantic segmentation by utilizing bi-directional learning with self-supervised learning functions. ADPL outperforms most models in domain adaptation by introducing two interactive single-domain adaptation paths aligned in the source and target domains to promote pseudo-labeling. The models mentioned are representative of domain adaptation and were therefore selected for comparison.

From Table 4, it can be found that the GAN-DANet proposed in this paper is superior to the current domain adaptation algorithms. The MMD series of algorithms underperforms, compared to other domain adaptation algorithms, and only surpasses fully supervised learning algorithms. This is primarily due to MMD being a non-parametric method that makes fewer assumptions about the sample distribution, rendering it more sensitive to distributional changes. In this study, the presence of shadows increased the disparity between the sample distributions in the source and target domains, rendering MMD ineffective at accurately measuring the differences, which led to poor performance. When only GANs are applied for domain adaptation, the transfer performance in crack segmentation remains inferior. Although the BDL and ADPL algorithms achieved superior transfer results, their performance exhibited significant variability between the source and target domains.

The GAN-DANet proposed in this study achieves excellent transfer results in both source and target domains by training in the new domain generated by CrackGAN, which can be well used for the transfer learning of cracks in shaded environments. This is primarily because GAN-DANet initially fuses the features of the source and target domains using CrackGAN to create a new dataset, which has more domain-invariant features. CrackSeg improves the model's ability to extract invariant features in the crack domain by connecting features of different resolutions in parallel and continuously fusing multiscale features to better utilize information at different scales.

Some advanced generative models such as StyleGAN (Karras et al., 2019) or Diffusion models (Ho et al., 2020), while state of the art in high-fidelity image generation, require significant computational resources and large-scale datasets for effective training. In contrast, the chosen methods achieve a balance between computational efficiency and segmentation accuracy, making them more suitable for the relatively small and domain-specific datasets used in this study.

Across the five independent runs on the fixed data split, the model demonstrated consistent performance in key metrics. This indicates that the model's performance is robust to variations in training dynamics caused by random initialization and optimization.

4.3.3 | Impact of different filters on the model

In the CrackSeg model, ELF was applied to enhance the high-frequency information such as the edges and details of the cracks and filter out the low-frequency information such as shadows. The performance gap between domain A and domain B is expected, as the model is not trained on domain B. Despite this, the proposed filtering method demonstrates strong generalization capabilities, achieving consistent improvements in domain A and competitive results in domain B. To verify the validity of ELF, different high-pass filters were used as comparison experiment, and the results are shown in Table 5. The impact of high-pass filtering on transfer learning for crack segmentation in shaded environments is consistently effective. The ELF this research employed outperforms other methods in the target domain, whereas the Sobel filter demonstrates lower efficacy. In addition, the effect of trainable high-pass filters lies between the Sobel class and the Laplace class. This is attributable to this study's proposed ELF being a Laplace-based high-pass filter, which identifies edge information through the detection of second-order derivatives of luminance changes in images. In contrast, the Sobel filter identifies edges solely by the first-order derivatives of luminance changes and determines the direction of the edges. However, since Sobel is less sensitive to noise, it is not as good as Laplace-like filtering, which is more sensitive to noise. The results in Table 5 further demonstrate the effectiveness of ELF in CrackSeg.

4.3.4 | Visual comparison

This study tested GAN-DANet in the test set in domain B. The results are shown in Figure 7. From the figure, it can be seen that GAN-DANet still segments out the cracks in the darker-shaded environments. Its segmentation performance is very close to the ground truth. In addition, images of actual shadows were collected to evaluate the segmentation performance of GAN-DANet under real shadow conditions, with the results presented in Figure 8. It can be found that the HRNet obtained by training in the source domain can hardly distinguish shadows from cracks. GAN-DANet can accurately differentiate between shadows and cracks in real-world scenarios. Moreover, it also has high accuracy for targets that may cause misjudgment, such as oil spots on the pavement, which is better than the ADPL. Compared with the BDL model, GAN-DANet has higher detection accuracy for both wider and narrower cracks. Figure 8 further demonstrates the excellence of GAN-DANet for crack segmentation in shaded environments.

Despite relying on generated shadow datasets, the evaluation metrics and actual detection results demonstrate the generalization ability of the model. Specifically, the model achieves consistent performance across both synthetic shadow datasets and real-world test datasets, indicating its robustness to domain shifts.

While GAN-DANet demonstrates robust performance in shadowed environments, additional challenges remain in other real-world conditions, such as wet and slippery roads. In wet environments, the reduced contrast and noise introduced by water reflections make crack segmentation even more challenging than in shadowed conditions. Future studies will focus on extending the proposed framework to address these scenarios by generating realistic wet crack datasets and incorporating advanced noise suppression techniques into the model.

4.4 | Ablation studies and discussion

To evaluate the effectiveness of each subnetwork within the GAN-DANet framework, ablation studies were conducted, with the results detailed in Table 6. When CrackGAN fuses data from both the source and target

TABLE 5 Comparison with different filters.

	Domain A			Domain B		
Model	**mIoU**	**mPA**	**Accuracy**	**mIoU**	**mPA**	**Accuracy**
1	74.62	80.67	96.89	74.83	80.81	96.91
2	73.95	79.69	97.09	73.66	79.27	96.78
3	74.13	79.86	97.12	73.84	79.75	96.81
4	74.78	80.58	97.19	74.22	79.44	96.9
5	74.91	80.73	97.2	74.37	79.78	96.91
This study	**75.06**	**80.77**	**97.23**	**75.03**	80.74	**96.95**

Note: 1–Laplace, 2–horizontal Sobel, 3–vertical Sobel, 4–high pass filtering, 5–trainable high-pass filtering. Bold values refer that proposed model is superior in these metrics.

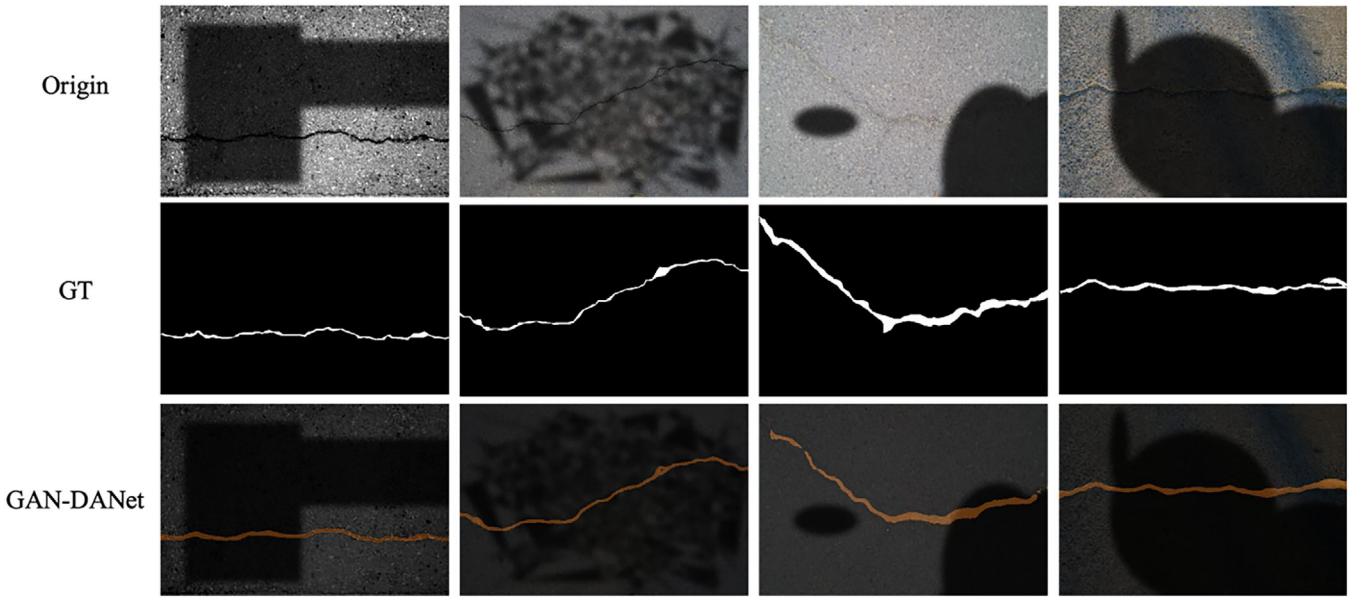

FIGURE 7 Results of GAN-DANet in domain B.

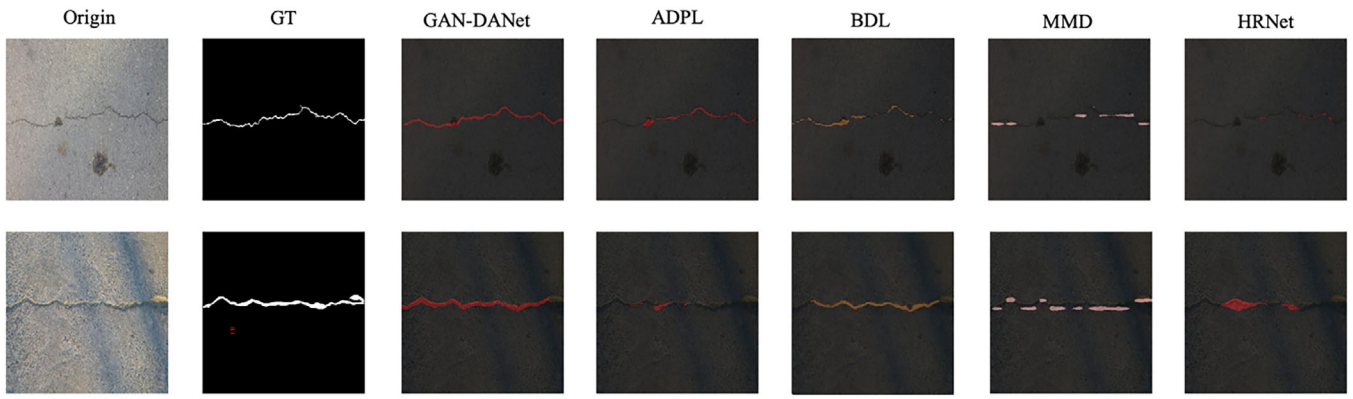

FIGURE 8 Results of different domain adaptation algorithms for detecting cracks in real shaded environments.

domains to create a new dataset for training, the model's performance diminishes in the source domain. Conversely, there is a significant improvement in the target domain, where the mIoU value increases from 57.87% to 73.96%. While the segmentation model with ELF has greatly improved the performance of crack segmentation on both source and target domains, its performance in the target domain further proves that the ELF can improve the

TABLE 6 Ablation Studies.

Model	Domain A			Domain B		
	mIoU	mPA	Accuracy	mIoU	mPA	Accuracy
Baseline	77.51	85.08	97.71	57.87	66.86	94.21
+N1	74.28	79.41	97.11	73.96	79.24	96.77
+N2	78.02	85.73	97.89	63.94	68.16	95.68
+N1+N2	75.06	80.77	97.23	75.03	80.74	96.95

Note: N1 refers to CrackGAN, which indicates that the segmentation model was trained on the generated dataset. N2 refers to CrackSeg, which indicates that the ELF-based segmentation model was used to train the model in the source domain.

model's ability to distinguish between cracks and shadow edges. By comparing the results of N1 and N2, it can be found that CrackGAN lies in improving the model's ability of domain adaptation, while CrackSeg lies in filtering the shadow features and enhancing the segmentation model's ability of extracting crack edges. The combination of the two further improves the cracks segmentation performance of the model in shaded environments, with the mIoU value improving from 57.87% to 75.03%.

5 | CONCLUSION

This research focuses on domain adaptation techniques for high-precision segmentation algorithms of cracks in shaded environments, which is particularly important for intelligent transportation detection systems. Based on this, this research proposes a two-stage model GAN-DANet for crack segmentation in shaded environments. GAN-DANet initially merges the unshaded and shaded crack datasets using the sub-network CrackGAN, which generates domains enriched with invariant features specific to the crack domain. Meanwhile, another sub-network CrackSeg enhances the edge and texture features of cracks during feature extraction by integrating the enhanced Laplace filtering and filtering the shadow information. GAN-DANet can improve the transfer accuracy of crack segmentation from 57.87% to 75.03%. Comparison with the segmentation results of state-of-the-art domain adaptation algorithms in real shadow images further demonstrates the performance of GAN-DANet.

In the future, more sophisticated supervised machine learning or classification algorithms could be explored as alternative approaches to address the pavement defects detection or segmentation in complex environments. For example, methods such as neural dynamic classification algorithms, StyleGAN, Diffusion, and the finite element machine for fast learning could offer new perspectives and enhance model performance. Additionally, self-supervised learning techniques may present an opportunity to leverage unlabeled data for better feature extraction and model generalization. Incorporating these advanced techniques could provide more robust and adaptable solutions for pavement defects detection or structural health monitoring in complex environments. In addition, this study acknowledges that the generated shadows may not fully replicate the complexity of all real-world shadows. Some advanced shadow generation techniques, such as using physics-based rendering, and more real-world shadowed datasets will be collected to further improve model robustness.

ACKNOWLEDGMENTS

The research is partly supported by the New Faculty Start-up Fund from the City University of Hong Kong with Grant Number 9610612, partly supported by the Research Matching Grant Scheme with Grant Number 9229141, partly supported by the CityU Strategic Interdisciplinary Research Grant with Grant Number 7020076.

REFERENCES

Alam, K. M. R., Siddique, N., & Adeli, H. (2020). A dynamic ensemble learning algorithm for neural networks. *Neural Computing and Applications*, *32*(12), 8675–8690. https://doi.org/10.1007/s00521-019-04359-7

Çelik, F., & König, M. (2022). A sigmoid-optimized encoder–decoder network for crack segmentation with copy-edit-paste transfer learning. *Computer-Aided Civil and Infrastructure Engineering*, *37*(14), 1875–1890.

Cheng, Y., Wei, F., Bao, J., Chen, D., & Zhang, W. (2023). ADPL: Adaptive dual path learning for domain adaptation of semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(8), 9339–9356.

Chun, P.-J., & Kikuta, T. (2024). Self-training with Bayesian neural networks and spatial priors for unsupervised domain adaptation in crack segmentation. *Computer-Aided Civil and Infrastructure Engineering*, *39*(17), 2642–2661.

Dais, D., Bal, I. E., Smyrou, E., & Sarhosis, V. (2021). Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning. *Automation in Construction*, *125*, 103606.

Nardin, A. D., Zottin, S., Piciarelli, C., Colombi, E., & Foresti, G. L. (2023). Few-shot pixel-precise document layout segmentation via dynamic instance generation and local thresholding. *International Journal of Neural Systems*, *33*(10), 2350052.

Díaz-Francés, J. Á., Fernández-Rodríguez, J. D., Thurnhofer-Hemsi, K., & López-Rubio, E. (2024). Semi-supervised semantic image segmentation by deep diffusion models and generative adversarial networks. *International Journal of Neural Systems*, *34*(11), 2450057.

Fan, L., Li, S., Li, Y., Li, B., Cao, D., & Wang, F.-Y. (2023). Pavement cracks coupled with shadows: A new shadow-crack dataset and a shadow-removal-oriented crack detection approach. *IEEE/CAA Journal of Automatica Sinica*, *10*(7), 1593–1607.

Fan, X., Cao, P., Shi, P., Chen, X., Zhou, X., & Gong, Q. (2022). An underwater dam crack image segmentation method based on multi-level adversarial transfer learning. *Neurocomputing*, *505*, 19–29.

Fu, J., Peng, H., Li, B., Liu, Z., Lugu, R., Wang, J., & Ramírez-de Arellano, A. (2024). Multitask adversarial networks based on extensive nonlinear spiking neuron models. *International Journal of Neural Systems*, *34*(6), 2450032.

Gao, Y., Zhai, P., & Mosalam, K. M. (2021). Balanced semisupervised generative adversarial network for damage assessment from low-data imbalanced-class regime. *Computer-Aided Civil and Infrastructure Engineering*, *36*(9), 1094–1113.

García-Aguilar, I., García-González, J., Luque-Baena, R. M., López-Rubio, E., & Domínguez, E. (2023). Optimized instance segmentation by super-resolution and maximal clique generation. *Integrated Computer-Aided Engineering*, *30*(3), 243–256.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. C., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144.

He, X., Tang, Z., Deng, Y., Zhou, G., Wang, Y., & Li, L. (2023). UAV-based road crack object-detection algorithm. *Automation in Construction*, *154*, 105014.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, *574*, 6840–6851.

Hoang, N.-D., & Nguyen, Q.-L. (2019). A novel method for asphalt pavement crack classification based on image processing and machine learning. *Engineering with Computers*, *35*, 487–498.

Hu, J., Yu, C., Yi, Z., & Zhang, H. (2023). Enhancing robustness of medical image segmentation model with neural memory ordinary differential equation. *International Journal of Neural Systems*, *33*(12), 2350060.

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(12), 4217–4228.

Katsigiannis, S., Seyedzadeh, S., Agapiou, A., & Ramzan, N. (2023). Deep learning for crack detection on masonry façades using limited data and transfer learning. *Journal of Building Engineering*, *76*, 107105.

Leng, J., Zhu, J., Yan, Y., Yu, X., Liu, M., Lou, Y., Liu, Y., Gao, L., Sun, Y., He, T., Yang, Q., Feng, C., Wang, D., Zhang, Y., Xu, Q., & Xu, F. (2024). Multilevel laser-induced pain measurement with Wasserstein generative adversarial network-gradient penalty model. *International Journal of Neural Systems*, *34*(1), 2350067.

Li, Y., Che, P., Liu, C., Wu, D., & Du, Y. (2021). Cross-scene pavement distress detection by a novel transfer learning framework. *Computer-Aided Civil and Infrastructure Engineering*, *36*(11), 1398–1415.

Li, Y., Yuan, L., & Vasconcelos, N. (2019). Bidirectional learning for domain adaptation of semantic segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA (pp. 6936–6945).

Liu, J., Yang, X., Lau, S., Wang, X., Luo, S., Lee, V. C.-S., & Ding, L. (2020). Automated pavement crack detection and segmentation based on two-step convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering*, *35*(11), 1291–1305.

Liu, Y., Yao, J., Lu, X., Xie, R., & Li, L. (2019). DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing*, *338*, 139–153.

Ma, S., Song, K., Niu, M., Tian, H., Wang, Y., & Yan, Y. (2023). Shape-consistent one-shot unsupervised domain adaptation for rail surface defect segmentation. *IEEE Transactions on Industrial Informatics*, *19*(9), 9667–9679.

Maeda, H., Kashiyama, T., Sekimoto, Y., Seto, T., & Omata, H. (2021). Generative adversarial network for road damage detection. *Computer-Aided Civil and Infrastructure Engineering*, *36*(1), 47–60.

Meng, S., Gao, Z., Zhou, Y., He, B., & Djerrad, A. (2023). Real-time automatic crack detection method based on drone. *Computer-Aided Civil and Infrastructure Engineering*, *38*(7), 849–872.

Ottoni, A. L. C., Souza, A. M., & Novo, M. S. (2023). Automated hyperparameter tuning for crack image classification with deep learning. *Soft Computing*, *27*(23), 18383–18402.

Pereira, D. R., Piteri, M. A., Souza, A. N., Papa, J. P., & Adeli, H. (2020). FEMA: A finite element machine for fast learning. *Neural Computing and Applications*, *32*, 6393–6404. https://doi.org/10.1007/s00521-019-04146-4

Rafiei, M. H., & Adeli, H. (2017). A new neural dynamic classification algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(12), 3074–3083. https://doi.org/10.1109/TNNLS.2017.2682102

Rafiei, M. H., Gauthier, L. V., Adeli, H., & Takabi, D. (2022). Self-supervised learning for electroencephalography. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(2), 1457–1471. https://doi.org/10.1109/TNNLS.2022.3190448

Rafiei, M. H., Gauthier, L. V., Adeli, H., & Takabi, D. (2024). Self-supervised learning for near-wild cognitive workload estimation. *Journal of Medical Systems*, *48*(1), 107. https://doi.org/10.1007/s10916-024-02122-7

Shi, Y., Cui, L., Qi, Z., Meng, F., & Chen, Z. (2016). Automatic road crack detection using random structured forests. *IEEE Transactions on Intelligent Transportation Systems*, *17*(12), 3434–3445.

Shim, S. (2024). Self-training approach for crack detection using synthesized crack images based on conditional generative adversarial network. *Computer-Aided Civil and Infrastructure Engineering*, *39*(7), 1019–1041.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., & Xiao, B. (2020). Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(10), 3349–3364.

Wang, J., Zhang, L., & Zhang, Y. (2023). Mixture 2D convolutions for 3D medical image segmentation. *International Journal of Neural Systems*, *33*(1), 2250059.

Xue, Y., Zhang, Y., & Neri, F. (2023). A method based on evolutionary algorithms and channel attention mechanism to enhance cycle generative adversarial network performance for image translation. *International Journal of Neural Systems*, *33*(5), 2350026.

Yan, H., Ding, Y., Li, P., Wang, Q., Xu, Y., & Zuo, W. (2017). Mind the class weight bias: Weighted maximum mean discrepancy for

unsupervised domain adaptation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI (pp. 2272–2281).

Yang, F., Zhang, L., Yu, S., Prokhorov, D., Mei, X., & Ling, H. (2019). Feature pyramid and hierarchical boosting network for pavement crack detection. *IEEE Transactions on Intelligent Transportation Systems*, *21*(4), 1525–1535.

Yang, H., Huyan, J., Ma, T., Song, Y., & Han, C. (2024). A novel applicable shadow resistant neural network model for high efficiency grid level pavement crack detection. *IEEE Transactions on Artificial Intelligence*, *5*(9), 4535–4549.

Yang, Q., Shi, W., Chen, J., & Lin, W. (2020). Deep convolution neural network-based transfer learning method for civil infrastructure crack detection. *Automation in Construction*, *116*, 103199.

Zhang, F., Tian, M., Li, Z., Xu, B., Lu, Q., Gao, C., & Sang, N. (2024). Lookup table meets local Laplacian filter: Pyramid reconstruction network for tone mapping. *Proceedings of the 37th Conference on Neural Information Processing Systems*, New Orleans, LA (pp. 57558–57569).

Zhang, H., Qian, Z., Zhou, W., Min, Y., & Liu, P. (2024). A controllable generative model for generating pavement crack images in complex scenes. *Computer-Aided Civil and Infrastructure Engineering*, *39*(12), 1795–1810.

Zhang, T., Wang, D., Mullins, A., & Lu, Y. (2023). Integrated APC-GAN and AttUNet framework for automated pavement crack pixel-level segmentation: A new solution to small training datasets. *IEEE Transactions on Intelligent Transportation Systems*, *24*(4), 4474–4481.

Zhang, Y., & Liu, C. (2024a). Crack segmentation using discrete cosine transform in shadow environments. *Automation in Construction*, *166*, 105646.

Zhang, Y., & Liu, C. (2024b). Network for robust and high-accuracy pavement crack segmentation. *Automation in Construction*, *162*, 105375.

Zhang, Y., & Liu, C. (2024c). Real-time pavement damage detection with damage shape adaptation. *IEEE Transactions on Intelligent Transportation Systems*, *25*(11), 18954–18963. https://doi.org/10.1109/TITS.2024.3416508

Zhang, Y., Ma, Z., Song, X., Wu, J., Liu, S., Chen, X., & Guo, X. (2021). Road surface defects detection based on IMU sensor. *IEEE Sensors Journal*, *22*(3), 2711–2721.

Zhang, Y., & Zhang, L. (2024). A generative adversarial network approach for removing motion blur in the automatic detection of pavement cracks. *Computer-Aided Civil and Infrastructure Engineering*, *39*(22), 3412–3434.

Zhang, Y., Zuo, Z., Xu, X., Wu, J., Zhu, J., Zhang, H., Wang, J., & Tian, Y. (2022). Road damage detection using UAV images based on multi-level attention mechanism. *Automation in Construction*, *144*, 104613.

Zhong, C., Xiong, Y., Tang, W., & Guo, J. (2024). A stage-wise residual attention generation adversarial network for mandibular defect repairing and reconstruction. *International Journal of Neural Systems*, *34*(7), 2450033.

Zhu, G., Liu, J., Fan, Z., Yuan, D., Ma, P., Wang, M., Sheng, W., & Wang, K. C. (2024). A lightweight encoder–decoder network for automatic pavement crack detection. *Computer-Aided Civil and Infrastructure Engineering*, *39*(12), 1743–1765.

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy (pp. 2242–2251). https://doi.org/10.1109/ICCV.2017.244

Zhu, Q., & Ha, Q. P. (2022). A bidirectional self-rectifying network with Bayesian modeling for vision-based crack detection. *IEEE Transactions on Industrial Informatics*, *19*(3), 3017–3028.

How to cite this article: Zhang, Y., & Liu, C. (2025). Generative adversarial network based on domain adaptation for crack segmentation in shadow environments. *Computer-Aided Civil and Infrastructure Engineering*, 1–17. https://doi.org/10.1111/mice.13451