electronics

MDPI

Article

Generative AI-Driven Data Augmentation for Crack Detection in Physical Structures

Jinwook Kim, Joonho Seon, Soohyun Kim, Youngghyu Sun, Seongwoo Lee, Jeongho Kim, Byungsun Hwang and Jinyoung Kim *

Department of Electronic Convergence Engineering, Kwangwoon University, Seoul 01897, Republic of Korea; yoonlight12@kw.ac.kr (J.K.); dimlight13@kw.ac.kr (J.S.); kimsoogus@kw.ac.kr (S.K.); yakrkr@kw.ac.kr (Y.S.); swoo1467@kw.ac.kr (S.L.); jh980828@kw.ac.kr (J.K.); hbsun1225@kw.ac.kr (B.H.)
* Correspondence: jinyoung@kw.ac.kr; Tel.: +82-2-940-5567

Abstract: The accurate segmentation of cracks in structural materials is crucial for assessing the safety and durability of infrastructure. Although conventional segmentation models based on deep learning techniques have shown impressive detection capabilities in these tasks, their performance can be restricted by small amounts of training data. Data augmentation techniques have been proposed to mitigate the data availability issue; however, these systems often have limitations in texture diversity, scalability over multiple physical structures, and the need for manual annotation. In this paper, a novel generative artificial intelligence (GAI)-driven data augmentation framework is proposed to overcome these limitations by integrating a projected generative adversarial network (ProjectedGAN) and a multi-crack texture transfer generative adversarial network (MCT2GAN). Additionally, a novel metric is proposed to evaluate the quality of the generated data. The proposed method is evaluated using three datasets: the bridge crack library (BCL), DeepCrack, and Volker. From the simulation results, it is confirmed that the segmentation performance can be improved by the proposed method in terms of intersection over union (IoU) and Dice scores across three datasets.

Keywords: crack detection; data augmentation; generative artificial intelligence; generative adversarial network; semantic segmentation

check for updates

Citation: Kim, J.; Seon, S.; Kim, S.; Sun, Y.; Lee, S.; Kim, J.; Hwang, B.; Kim, J. Generative AI-Driven Data Augmentation for Crack Detection in Physical Structures. *Electronics* **2024**, *13*, 3905. https://doi.org/10.3390/electronics13193905

Academic Editors: Galina Ilieva, George A. Tsihrintzis

Received: 6 September 2024
Revised: 30 September 2024
Accepted: 1 October 2024
Published: 2 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Cracks are a common problem in physical structures and can cause the deterioration of their durability. The prompt detection and repair of cracks are essential for preventing accidents that may lead to structural collapse [1–3]. Regular crack inspection has been used to assess structural stability, but conventional visual inspection is labor-intensive and relies on the subjective judgment of the inspector. This reliance can lead to inconsistencies in the assessment process and compromise the structure's safety. Image processing methods based on thresholding [4] and edge detection [5] techniques have been proposed to address these inconsistencies. Nevertheless, the performance of these techniques can be degraded by unexpected noise similar to cracks. In addition, their dependency on predefined thresholding [4] and linear filtering [5] limits their generalization ability to new crack images.

Remarkable achievements in deep learning for computer vision have influenced its application to segmentation methodologies. Specifically, convolutional neural network (CNN)-based methods can more effectively identify crack patterns compared to previous image processing techniques thanks to their ability to approximate intricate nonlinear functions [6–8]. Among the CNN-based approaches, semantic segmentation has emerged as an important research area in crack detection due to its capability in detecting cracks and providing attribute information. Several CNN-based approaches have been proposed to perform crack segmentation [9–11]. Fully connected networks (FCNs) have been proposed

to extract information on the detailed attributes of cracks, such as width, height, and location [9]. Based on its skip connections and symmetric structure, U-Net can reduce false positives caused by unexpected noise [10]. Furthermore, DeepCrack has been proposed to address the discontinuity of crack detection by combining low-level and high-level features [11]. However, CNN-based crack detection methodologies are highly dependent on the quality and variety of the training data, which can limit their generalization in practical applications [12–15].

Because various crack shapes and surface characteristics (texture, rugometry, brightness, glow, etc.) can appear in real-world physical structures, collecting diverse images is crucial to enhance the generalization performance of CNN-based crack detection methods [16,17]. However, when capturing crack images of physical structures, the availability of datasets is often limited due to the cost of data collection. Generative artificial intelligence (GAI), specifically data augmentation techniques based on generative adversarial networks (GANs), has been proposed to address this data availability challenge. GAN-based approaches can complement the scarcity of label data thanks to their ability to generate plausible and diverse images [18]. A data augmentation method based on deep convolutional GAN (DCGAN) [19] has been proposed to augment small crack datasets. However, DCGANs may suffer instability issues because the discriminator converges faster than the generator. The automated pavement crack GANs (APC-GANs) method has been developed to address this convergence problem by incorporating Gaussian noise into the discriminator to reduce its convergence speed [20].

While GAN models have advantages, conventional methods [19,20] require manual annotation, which leads to additional costs. Therefore, the automation of manual annotation can be crucial to mitigate the costs. Researchers have proposed a method to transfer crack textures between different structures while preserving crack patterns using CycleGAN-based techniques to automate this process [21]. However, CycleGAN has limitations in changing geometric shapes, which may limit the texture diversity of generated crack images. A framework integration of DCGAN and pixel-to-pixel (Pix2Pix) has been proposed to address the diversity issue [22]. While the DCGAN-with-Pix2Pix method can be used to improve crack diversity, it remains dependent on human visual inspection for quality assurance. Therefore, there is a need to automate quality assurance inspection while increasing diversity.

In summary, the data availability issue has been mitigated using data augmentation in previous methods. However, manual annotation can incur costs for conventional data augmentation methods. While recent studies have been proposed to reduce the annotation cost, previous studies encountered challenges related to texture diversity, scalability over multiple physical structures, and the need for manual inspection. In this paper, a novel GAI-driven data augmentation framework is proposed to address these issues by incorporating projected GAN (ProjectedGAN) and multi-crack texture transfer GAN (MCT2GAN). The proposed framework can mitigate the constraints of texture diversity and scalability over multiple physical structures by employing ProjectedGAN and MCT2GAN, respectively, while reducing the costs of manual annotation via a novel evaluation process. The main contributions of this paper can be summarized as follows:

- A novel dataset augmentation framework is proposed to address the limitations of texture diversity and data availability issues in small crack datasets. The proposed framework operates in two stages: a label generator and a crack image generator. By implementing this framework to generate a diverse texture crack dataset, the segmentation accuracy in crack detection systems can be improved in environments with small amounts of available data.
- A modified loss function and ReMix method are integrated into StarGANv2; MCT2GAN is proposed to preserve the label image's crack shape and improve GAN's learning stability. The proposed GAN model that can preserve the crack shape can improve the segmentation performance in crack detection systems with small amounts of crack data.

- A novel performance metric called the "crack representation balance (CRB) score" is proposed to assess the representation of crack patterns in generated images by considering global and local perspectives. The structure of the cracks and the detailed patterns of each can be qualified with global and local perspectives, respectively. Therefore, this approach can provide a perspective to evaluate the quality of the generated crack segmentation dataset.

The rest of this paper is organized as follows: In Section 2, a dataset synthesis framework is proposed. In Section 3, the proposed framework is validated through experiments. Section 4 concludes the paper by discussing this study's findings and suggesting areas for future research.

2. Framework for Crack Data Augmentation

The process of generating synthetic crack datasets in the proposed dataset synthesis framework is illustrated in Figure 1. This generation process consists of three stages: label generation, crack generation, and quality assessment. In the initial stage, a set of synthetic crack label images is generated by the ProjectedGAN model. A feedback loop is applied to the output to reduce the need for manual evaluation by incorporating a discriminator. The synthesized images are used to train the MCT2GAN model. Then, a style encoder is used to adjust the crack pattern to match the surface texture so that the cracks match the unique properties of each texture. The quality of the generated images can be evaluated with performance metrics to ensure that the best images are selected for inclusion. Finally, the synthesized label images are combined with the corresponding crack images to generate a diverse dataset suitable for training and testing the semantic segmentation model.

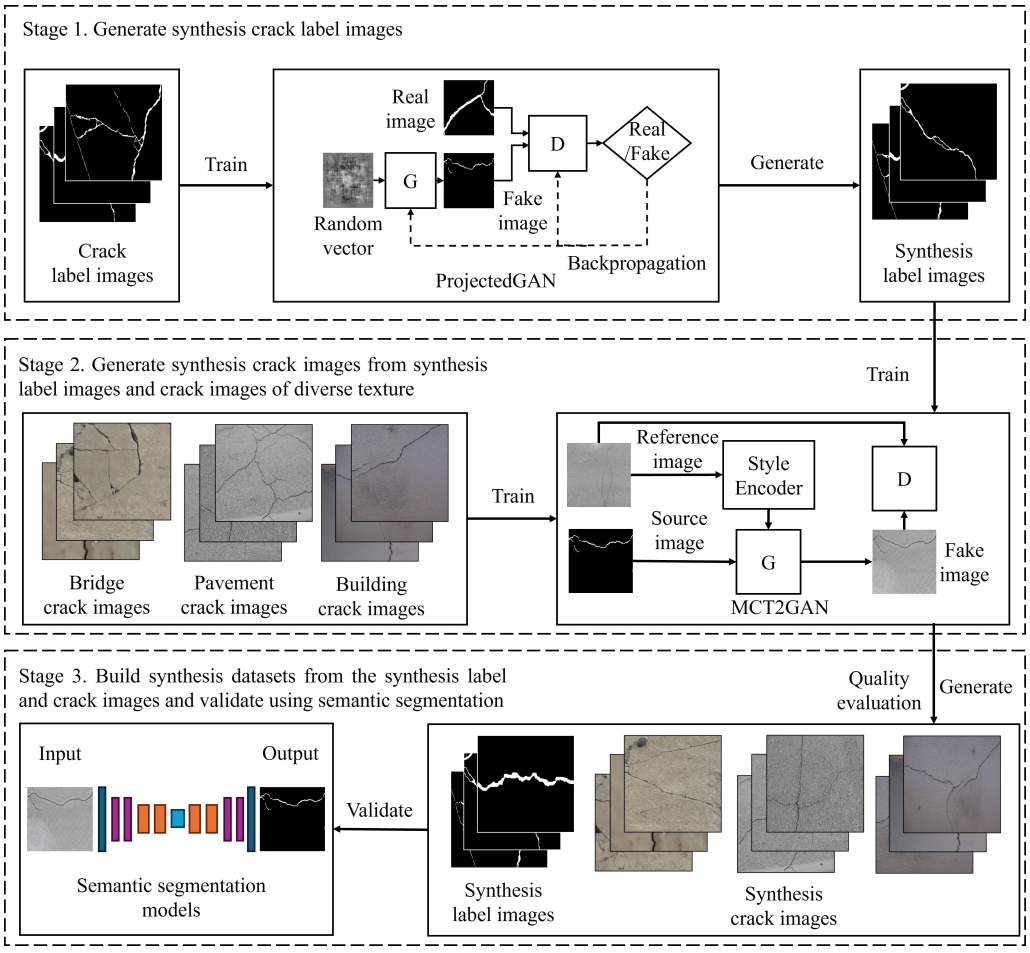

Figure 1. Process of generating a synthesized crack dataset in the proposed dataset synthesis framework.

2.1. Label Generator

ProjectedGAN [23] is a promising candidate for implementing a label generator, as it can generate diverse images. As shown in Figure 2, ProjectedGAN consists of a generator, a pre-trained feature network, a random projection, and multiscale discriminators. The generator G produces fake labeled images $G(z)$ from a random vector z. Features of a real label image x and a fake label image $G(z)$ are extracted by pre-trained feature extraction networks to disentangle causal generative factors. However, according to [23], the discriminator may not fully exploit the features extracted from the feature extraction networks. In Figure 2, random projection is employed to enhance the discriminator's capacity to fully utilize features extracted from the pre-trained networks, facilitating the mixing of features across channels and resolutions through 1×1 and 3×3 convolutions. Then, the multiscale discriminators D_l receive the projected features $P_l(\cdot)$ as inputs and determine whether they are real or fake. These multiscale discriminators provide feedback, encouraging the generator to create finely detailed images. Finally, the generator and discriminator weights are updated based on the feedback from the discriminators. The ProjectedGAN optimization process can be expressed as follows:

$$\min_{G} \max_{D_l} \sum_{l=1}^{4} \Big(\mathbb{E}_x[\log D_l(P_l(x))] + \mathbb{E}_z[\log(1 - D_l(P_l(G(z))))] \Big), \tag{1}$$

where the index l from 1 to 4 is used to represent the projected features P_l and discriminators D_l. Only D_l and G are updated during optimization, while P_l remains unchanged. P_l consists of pre-trained feature networks, cross-channel mixing (CCM), and cross-scale mixing (CSM).

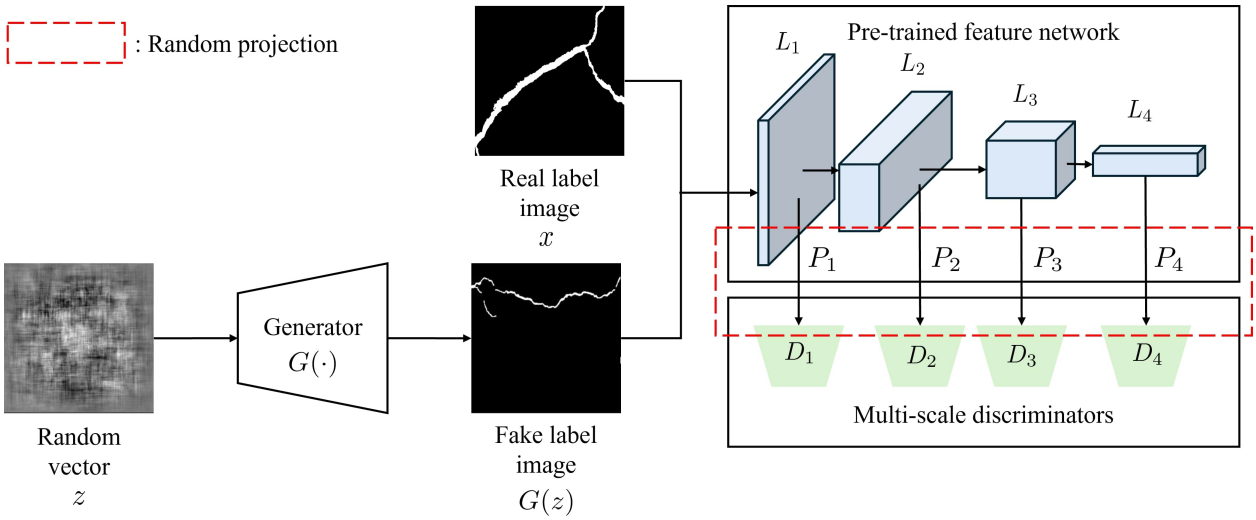

Figure 2. Structure of ProjectedGAN.

2.2. Crack Generator with Various Textures

MCT2GAN is an image-to-image translation model that can train multiple domains with a single generator, reducing the need for a re-training process for different data types. This ability to handle multiple domains can provide it with an advantage in scalability over multiple physical structures compared to the other GAN models trained on a specific structure. As shown in Figure 3, the MCT2GAN model is used to produce a wide range of textural crack images using both synthetic label and real crack datasets. Initially, the source and reference images are selected for translation. The reference image is passed through a style encoder, generating a style code, which the generator uses to transform the source image. Finally, the discriminator determines whether the transformed image is real or fake.

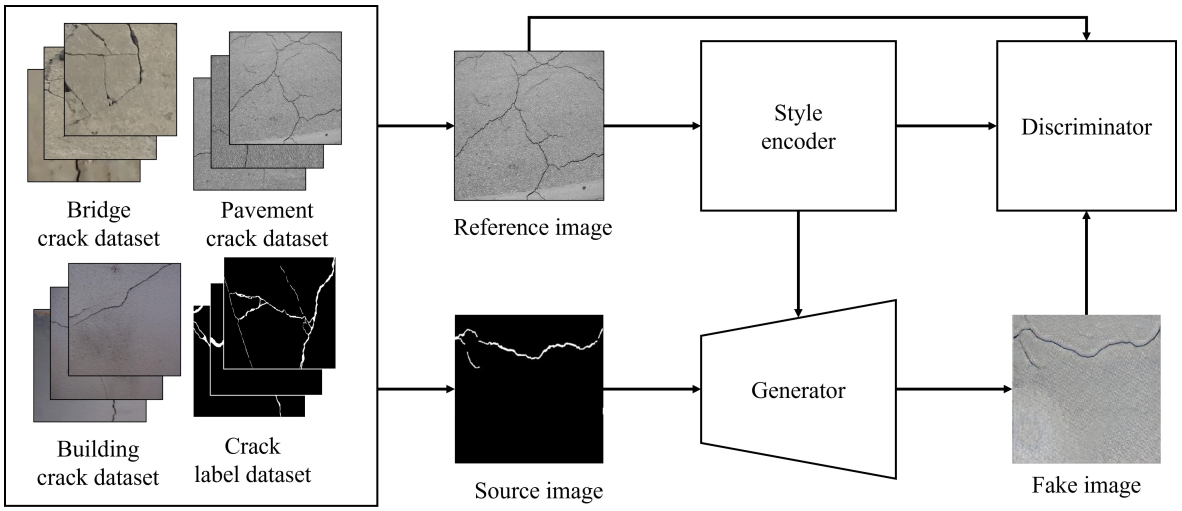

Figure 3. Structure of MCT2GAN.

The loss function of MCT2GAN consists of adversarial loss, style reconstruction, diversity-sensitive loss, and cycle consistency loss. The generator is designed to perform bi-directions, from image to label and from label to image. However, since the focus is on translating images from labels, the description will be limited to this process. In this context, the source images $\mathbf{x}_{src}$ and the reference images $\mathbf{x}_{ref}$ are assumed to be the label and crack images, respectively. y is the domain vector for each structure and consists of a building, pavement, and bridge.

In adversarial loss, the generator learns to generate a realistic crack images $G(\mathbf{x}_{src}, \tilde{\mathbf{s}})$ from the label images $\mathbf{x}_{src}$ and style code $\tilde{\mathbf{s}}$, making it indistinguishable from a crack image. Hinge loss [24] is used to implement adversarial loss to improve the training stability of MCT2GAN. The equation is represented as

$$\mathcal{L}_{adv,D} = \mathbb{E}_{\mathbf{x}_{src},y}[\max(0, 1 - D_y(\mathbf{x}_{src}))] + \mathbb{E}_{\mathbf{x}_{src},\mathbf{x}_{ref},\tilde{y}}[\max(0, 1 + D_{\tilde{y}}(G(\mathbf{x}_{src}, \tilde{\mathbf{s}})))], \quad (2)$$

$$\mathcal{L}_{adv,G} = \mathbb{E}_{\mathbf{x}_{src},\mathbf{x}_{ref},\tilde{y}}[D_{\tilde{y}}(G(\mathbf{x}_{src}, \tilde{\mathbf{s}}))], \quad (3)$$

where $\mathbf{x}_{src}$ and $\mathbf{x}_{ref}$ denote sets of label images and crack images, respectively, $D_y(\cdot)$ is the output of discriminator corresponding to domain y, and $\tilde{\mathbf{s}} = E(\mathbf{x}_{ref})$ is style code extracted by style encoder E from crack image $\mathbf{x}_{ref}$ corresponding to y.

The generator's objective in style reconstruction loss is to accurately capture the style of the structural domain $\tilde{y}$. The loss function of style reconstruction is expressed as

$$\mathcal{L}_{sty} = \mathbb{E}_{\mathbf{x}_{src},\tilde{y},\mathbf{x}_{ref}}[\|\tilde{\mathbf{s}} - E_{\tilde{y}}(G(\mathbf{x}_{src}, \tilde{\mathbf{s}}))\|_1], \quad (4)$$

where $E_{\tilde{y}}$ is the style encoder for the target domain $\tilde{y}$. During training, the generator can learn the style information by minimizing the difference between the encoded feature and the style code from the generated data.

The diversity-sensitive loss enables the generator to produce various texture crack images. This loss function encourages the generator to produce diverse outputs when provided with different style codes, even if the input label image is unchanged. The loss function is formulated as

$$\mathcal{L}_{ds} = \mathbb{E}_{\mathbf{x},\tilde{y},\mathbf{x}_{ref_1},\mathbf{x}_{ref_2}}[\|G(\mathbf{x}_{src}, \tilde{\mathbf{s}}_1) - G(\mathbf{x}_{src}, \tilde{\mathbf{s}}_2)\|_1], \quad (5)$$

where $\tilde{\mathbf{s}}_1 = E(\mathbf{x}_{ref_1})$ and $\tilde{\mathbf{s}}_2 = E(\mathbf{x}_{ref_2})$ are the style codes extracted from the reference images $\mathbf{x}_{ref_1}$ and $\mathbf{x}_{ref_2}$, respectively. By incorporating this loss, the generator is guided to avoid producing similar outputs and instead generates diverse texture patterns.

Cycle consistency loss is used to preserve the characteristics of the label image. In this study, the label image and the crack image must be paired in terms of the shape and localization of the crack to be utilized in semantic segmentation. Therefore, this loss plays an important role in generating an accurate crack image corresponding to the label image. This loss is defined as

$$\mathcal{L}_{cyc} = \mathbb{E}_{\mathbf{x},y,\tilde{y},\mathbf{z}}[\|\mathbf{x}_{src} - G(G(\mathbf{x}_{src}, \tilde{\mathbf{s}}), \hat{\mathbf{s}})\|_1], \tag{6}$$

where $\hat{\mathbf{s}} = E_y(\mathbf{x})$ represents the style code of the label crack image estimated by the style encoder.

By combining $\mathcal{L}_{sty}$, $\mathcal{L}_{ds}$, and $\mathcal{L}_{cyc}$ with their corresponding hyperparameters λ_{sty}, λ_{ds}, and λ_{cyc}, respectively, the loss function of MCT2GAN is formulated as follows:

$$\min_{G,E} \max_{D} \mathcal{L}_{adv} + \lambda_{sty}\mathcal{L}_{sty} - \lambda_{ds}\mathcal{L}_{ds} + \lambda_{cyc}\mathcal{L}_{cyc}. \tag{7}$$

This objective function enables the generator to produce realistic and diverse crack images across multiple domains.

2.3. Quality Evaluation for Generated Data

The Fréchet inception distance (FID) [25] is used to evaluate the fidelity of generated image data through the Fréchet distance between the fake image distribution P_f and the real image distribution P_r using InceptionNet. The formula of FID is as follows:

$$\mathrm{FID}(P_r, P_f) = \|\mu_r - \mu_f\|_2^2 + \mathrm{Tr}\Big(C_r + C_f - 2(C_r C_f)^{1/2}\Big), \tag{8}$$

where the mean vectors are denoted as μ_r and μ_f, and the covariance matrices are denoted as C_r and C_f for real and fake images, respectively. $\mathrm{Tr}(\cdot)$ means the sum of the diagonal elements of a matrix.

Quantifying whether the generated image represents the crack pattern in a specific label image can be used to evaluate the generator's performance. Evaluating the generator consists of three steps: generation, prediction, and comparison. First, the generator generates a crack image from the label image, a reference image. Then, image cracks are predicted by using the FCN model. The quality of the generated images can be evaluated by comparing the predicted images with the label images of the generated crack image and calculating the Dice, Hausdorff distance (HD) [26], and crack representation balance (CRB) scores. Lastly, the evaluated images are used to augment the original dataset. The intersection over union (IoU) and Dice scores are used to evaluate the augmented crack datasets' segmentation performance. The IoU and Dice scores are calculated as follows:

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{9}$$

$$\mathrm{Dice} = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{10}$$

where TP, FP, and FN are the true positive, false positive, and false negative calculated between the predicted image and the label image.

The process for assessing the quality of the generated crack images is depicted in Figure 4. Herein, the Dice score is used to evaluate the degree to which the generated crack represents the size of the label crack. The HD score can be used to evaluate the accuracy of the generated crack image in representing the shape of the labeled crack. The HD score is expressed as

$$\mathrm{HD}(A, B) = \max[h(A, B), h(B, A)], \tag{11}$$

where A and B are the sets of label images and generated crack images, respectively. $h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$ is the directed Hausdorff distance. a and b are points of the sets.

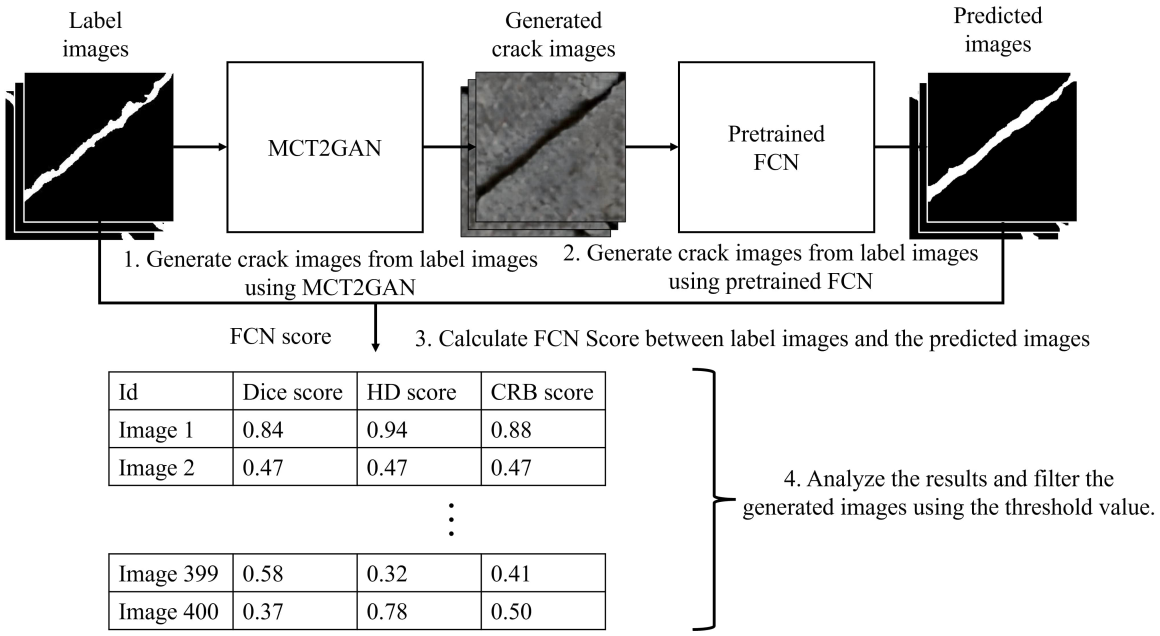

Figure 4. Process of quality evaluation.

The Dice and HD scores can measure the ability of a model to represent the size and patterns of cracks, respectively. These two metrics need to be considered simultaneously to evaluate crack representation. Therefore, the CRB score is proposed to consider the performance of both metrics simultaneously through the harmonic mean of the two metrics. The CRB score can be described using the following formulas:

$$\text{CRB-score} = 2 \times \frac{\text{Dice} \times \text{HD}}{\text{Dice} + \text{HD}}. \tag{12}$$

3. Simulation Results

3.1. Simulation Settings

The proposed framework was validated using three datasets: the Bridge Crack Library (BCL), DeepCrack, and Volker. These datasets were selected due to their diversity in crack types and environmental conditions. The BCL dataset consists of 11,000 images with a resolution of 256 × 256, which are collected from concrete bridges [27]. The number of images was selected to total 500 via random sampling to prevent generating crack images with a biased texture. The DeepCrack dataset contains 537 crack images with a resolution of 544 × 384, obtained using a line-array camera on the surface of the concrete and asphalt pavement [11]. The Volker dataset consists of 427 images with a resolution of 512 × 512 [28] in concrete building structures. All images were resized to a 256 × 256 resolution for consistency in the input dimensions. In this paper, datasets of small size are employed to exploit the advantage of the augmentation method. It is expected that the impact of augmentation will become smaller with decreasing data size. In other words, the gap in performance enhancement by the augmentation method may be decreased in large datasets. Therefore, small datasets could be more suitable for clearly demonstrating the impact of augmentation [29]. An overview of the datasets is given in Table 1.

Table 1. Overview of datasets.

Dataset	Structure	Materials	Dataset Size	Image Size (H × W)
BCL [27]	Bridge	Concrete	500	256 × 256
DeepCrack [11]	Pavement	Concrete and asphalt	537	544 × 384
Volker [28]	Building	Concrete	427	512 × 512

The label and crack generators were implemented using PyTorch 2.0.1, CUDA 11.7, and Python 3.10.11 on Ubuntu 20.04.5. To meet the computational requirements of the GAN model, two NVIDIA GeForce RTX 4090 GPUs were used for the simulations.

For the hyperparameter settings of ProjectedGAN, the number of iterations was empirically set to 20,000. The optimizer was configured as Adam, with a $\beta 1$ of 0.5 and a $\beta 2$ of 0.999. The batch size was set to 8, balancing GPU memory usage and training stability. The learning rate was set to 0.0002, and to smooth the learning process, an exponential moving average (EMA) with a decay rate of 0.999 was applied.

For MCT2GAN's hyperparameter settings, the weights for R1 regularization, cyclic consistency loss, style reconstruction loss, and diversity-sensitive losses were set to 1, 1, 10, and 2, respectively. Adam was used to optimize the generator and discriminator with a $\beta 1$ of 0 and a $\beta 2$ of 0.99. The batch size was set to 8. Both the learning rate and weight decay were set to 0.0001. EMA was applied with a decay rate of 0.999. The ReMix method [30] was used to prevent overfitting in the training process. The training iteration was empirically set to stop at 20,000.

3.2. Performance Evaluation and Discussion

The augmentation and stylization capabilities of the proposed model were evaluated by comparing them with conventional methods for data augmentation [21,22]. CycleGAN and Pix2Pix were selected as representative methods among the existing augmentation methods. Since CycleGAN and Pix2Pix cannot perform texture transfer, these models were trained using individual datasets, while MCT2GAN was trained with multiple datasets simultaneously.

The texture diversity and quality of the images generated by each GAN model were evaluated by visualizing the crack images in Figure 5. The images generated by CycleGAN can preserve crack patterns in the visual evaluation, but cannot describe detailed textures. This limitation may be due to the loss function of CycleGAN, which only preserves the crack patterns. The images generated by Pix2Pix can preserve both textures and crack patterns. However, it was confirmed that Px2Pix cannot represent various texture styles because the generated images tend to have similar textures. As shown in Figure 5, the proposed model can produce diverse texture crack images compared with the other models. In addition, it was confirmed that the crack patterns generated by the proposed model have a higher resolution than those generated by CycleGAN and Pix2Pix. Consequently, these results imply that the proposed model can generate more diverse textures and higher-quality crack images than comparative models.

A comparison based on FID, described in Table 2, was performed to compare the quality of images generated by conventional GAN-based and the proposed models. In Table 2, it is confirmed that the MCT2GAN model shows superior FID values compared with the other models on all datasets. Typically, it is known that learning multiple textures can degrade performance. However, the simulation results, which show that the FID scores of MCT2GAN obtained after learning three textures were higher than those of the comparative models, may imply that the proposed model can learn complex distributions in multiple structures.

Table 2. FID of GAN-based methods.

Method	BCL	DeepCrack	Volker
CycleGAN [21]	212.46	245.79	240.05
Pix2Pix [22]	101.02	142.64	139.33
MCT2GAN	**86.17**	**132.68**	**126.41**

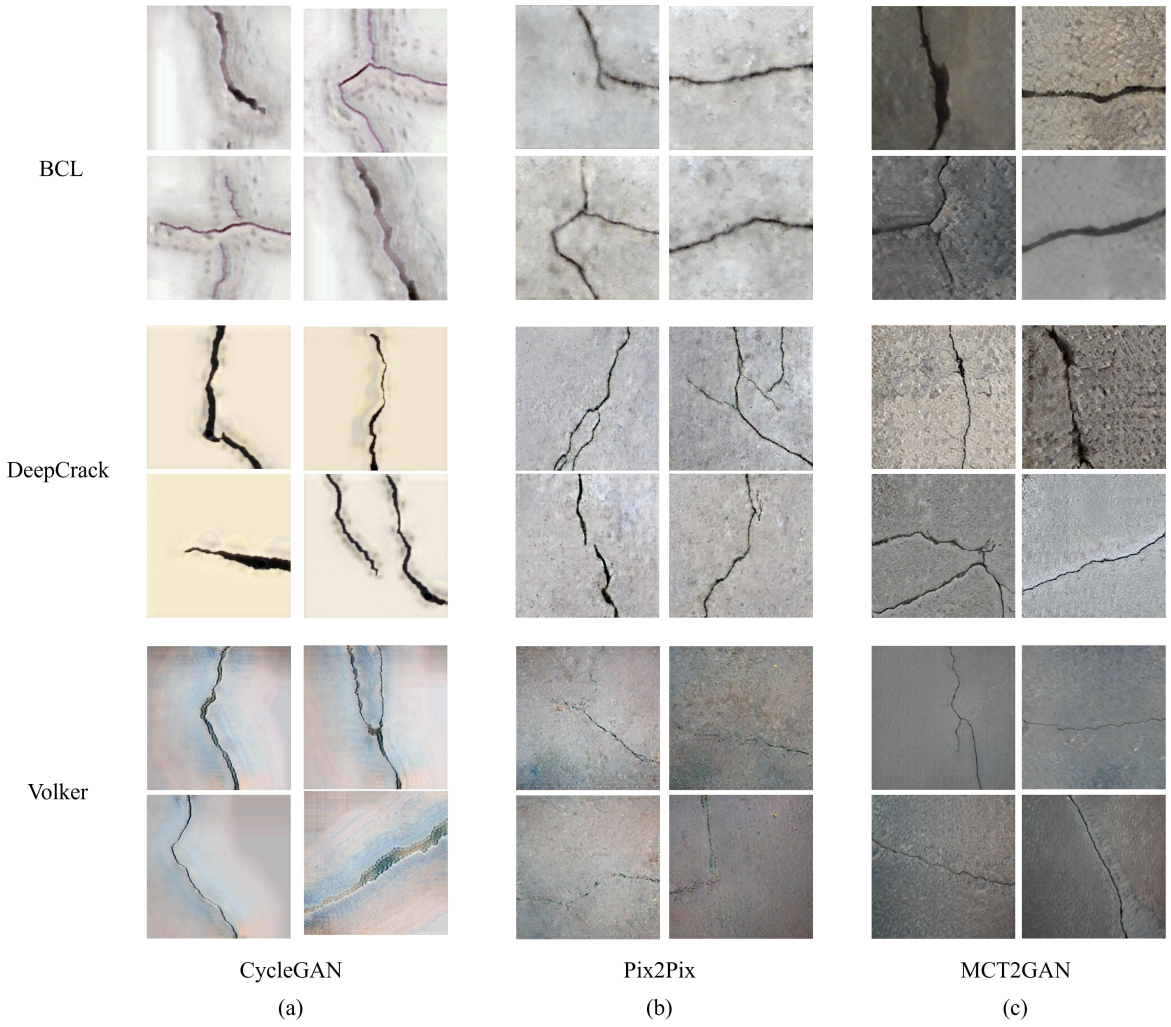

Figure 5. Crack images generated from GAN-based methods.

The quality metrics and images generated by the crack generator are described in Figures 6 and 7. After 2000 iterations, the HD score increased, indicating that the shape of the generated crack can be closer to the shape of the label image. However, both the Dice and CRB scores decreased. As shown in Figure 7a, this change was confirmed to be due to a reduction in overlapping areas between the label crack and the synthesized crack due to the crack being generated in the wrong location compared to the image in 1000 iterations. As a result, the CRB score effectively penalizes the synthesized crack's localization error, reflecting the positional discrepancy despite the improved shape accuracy. The Dice score barely changed between 6000 and 7000 iterations, as illustrated in Figure 6, while the HD and the CRB scores increased. In the overlapped image at 6000 and 7000 iterations, as shown in Figure 7b, the region of blue color (FP) below the crack decreased from 7000. It was confirmed that the HD score penalizes this exaggerated appearance of cracks. In this case, the CRB score effectively captures this penalty, reflecting the shape correction and overall crack quality. The HD score was maintained from 17,000 to 18,000 iterations, as illustrated in Figure 6, but both the Dice and the CRB scores increased. As shown in Figure 7c, the reduction in the blue region could indicate that the size of the synthesized crack changed. Here, the CRB score captures these changes effectively, reflecting the synthesized crack's size and shape. Consequently, the CRB score can be employed to assess the overall quality of synthesized crack images by simultaneously evaluating the images' localization, size, and shape.

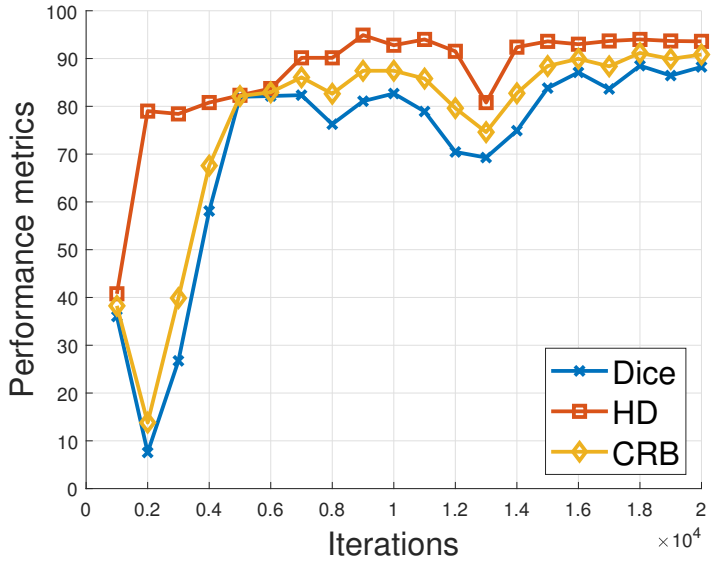

Figure 6. Performance metrics of generated crack image sample in training process.

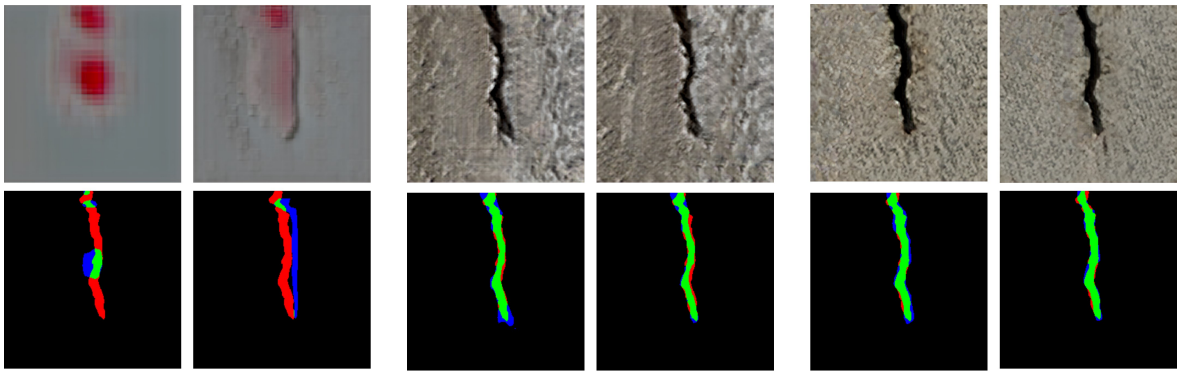

Figure 7. Samples of the generated crack images in the training process. The images in the first row are the synthesized crack images. The images in the second row are the overlapped images between the ground truth and the predicted mask of the synthesized image from the pre-trained FCN model.

An ablation study was performed to confirm the effectiveness of the label generator in data augmentation, and the results are summarized in Table 3. According to Table 3, increasing the cycle consistency loss λ_{cyc} can improve the Dice scores by preserving the shape of the crack. The ReMix method improved the Dice scores of BCL and DeepCrack, while Volker's performance suffered. By changing the adversarial loss from BCE to hinge loss, the variance in the Dice scores in the three datasets decreased. It was confirmed that hinge loss improves training stability by unscrewing the distribution of the three datasets. However, using hinge loss without the ReMix method decreased performance in all datasets. The model with hinge loss is typically trained to perform generalization for stable training. In environments with small amounts of data, hinge loss can lead to overfitting issues. ReMix, on the other hand, is a method that enhances the diversity of data based on augmentation techniques, thereby preventing overfitting. Therefore, upon integrating the ReMix method and hinge loss, they may act as complementary solutions to each other's weaknesses. In addition, a label generator can enhance the performance of MCT2GAN by increasing the amount of data.

Table 3. Ablation study of crack generator in terms of Dice score.

Method	BCL	DeepCrack	Volker
StarGANv2 [31]	39.66	36.67	33.86
StarGANv2 + λ_{cyc} 10	50.70	40.18	45.66
StarGANv2 + λ_{cyc} 10 + ReMix	58.79	45.86	37.21
StarGANv2 + λ_{cyc} 10 + Hinge Loss	35.81	26.77	33.33
StarGANv2 + λ_{cyc} 10 + ReMix + Hinge Loss (MCT2GAN)	60.36	52.57	51.71
MCT2GAN + Data of Label Generator (Proposed method)	**62.98**	**53.90**	**59.36**

The IoU and Dice scores of the original dataset and the dataset augmented by StarGANv2, MCT2GAN, and the proposed method are presented in Table 4. The results in Table 4 indicate that the U-Net trained on the dataset augmented by the proposed method outperformed the one trained on the original dataset and the dataset augmented by StarGANv2 and MCT2GAN regarding IoU and Dice score. For the BCL dataset, it was demonstrated that the IoU and Dice scores were up to 3.91% and 2.18% higher, respectively, than those of the original dataset. Then, for the DeepCrack dataset, it was demonstrated that the IoU and Dice scores improved by up to 5.00% and 3.23%, respectively. Lastly, for the Volker dataset, it was demonstrated that the IoU and Dice scores were up to 1.94% and 1.50% higher, respectively, than those of the original dataset. Based on the IoU and Dice scores, it is confirmed that the proposed method could enhance the crack segmentation task with small amounts of crack data.

Table 4. Crack dataset segmentation results.

Dataset	Original Dataset		Augmented Dataset					
			StarGANv2		MCT2GAN		Proposed Method	
	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
BCL	51.96	68.59	43.61	60.73	52.44	68.81	53.99	70.09
DeepCrack	68.37	81.04	58.83	74.08	69.87	82.26	71.79	83.66
Volker	42.78	59.64	33.34	50.01	38.40	55.49	43.61	60.54

4. Conclusions

In this paper, a novel generative artificial intelligence (GAI)-driven framework for effective data augmentation in environments with small amounts of data is proposed. By integrating a projected generative adversarial network (ProjectedGAN) and a multi-crack texture transfer generative adversarial network (MCT2GAN), the framework can address challenges related to texture diversity and scalability over multiple physical structures. In addition, the crack representation balance (CRB) score was introduced as a novel performance metric to evaluate the quality of generated crack images. From the simulation results, it was demonstrated that the proposed method can improve the performance of crack detection systems while reducing the costs associated with data collection and annotation on different datasets. The framework of the proposed method can be refined into an end-to-end system in future work to improve the efficiency of generating crack and label images. The style encoder can also be expanded into a local texture encoder to account for rugometry, brightness, and glow, thereby enhancing generalization to real-world physical structures. The proposed method may be applicable to medical datasets, which face challenges similar to those of crack datasets, including small amounts of data and a lack of diversity.

Author Contributions: Conceptualization, J.K. (Jinwook Kim), J.S., S.K. and S.L.; methodology, J.K. (Jinwook Kim), J.S., S.K. and S.L.; formal analysis, J.K. (Jinwook Kim), J.S., S.K., Y.S. and S.L.; investigation, J.K. (Jinwook Kim), B.H. and J.K. (Jeongho Kim); resources, J.K. (Jinwook Kim), B.H., and J.K. (Jeongho Kim); writing—original draft preparation, J.K. (Jinwook Kim), J.S., S.K., Y.S. and S.L.;

writing—review and editing, J.K. (Jinwook Kim), J.S., S.K., Y.S., S.L., B.H., J.K. (Jeongho Kim), and J.K. (Jinyoung Kim); visualization, J.K. (Jinwook Kim), B.H., J.K. (Jeongho Kim), and J.K. (Jinyoung Kim); supervision, J.K. (Jinyoung Kim); project administration, J.K. (Jinyoung Kim). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data is contained within the article.

Acknowledgments: This work was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-2020-0-01846) supervised by the IITP (Institute for Information and Communications Technology Planning and Evaluation), and in part by a Research Grant from Kwangwoon University in 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wardhana, K.; Hadipriono, F.C. Analysis of Recent Bridge Failures in the United States. *J. Perform. Constr. Facil.* **2003**, *17*, 144–150. [CrossRef]
2. Zhao, X.H.; Cheng, W.C.; Shen, J.S.; Arulrajah, A. Platform collapse incident of a power plant in Jiangxi, China. *Nat. Hazards* **2017**, *87*, 1259–1265. [CrossRef]
3. Tan, J.S.; Elbaz, K.; Wang, Z.F.; Shen, J.S.; Chen, J. Lessons learnt from bridge collapse: A view of sustainable management. *Sustainability* **2020**, *12*, 1205. [CrossRef]
4. Tang, J.; Gu, Y. Automatic crack detection and segmentation using a hybrid algorithm for road distress analysis. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 13–16 October 2013; pp. 3026–3030. [CrossRef]
5. Lim, R.S.; La, H.M.; Sheng, W. A robotic crack inspection and mapping system for bridge deck maintenance. *IEEE Trans. Autom. Sci. Eng.* **2014**, *11*, 367–378. [CrossRef]
6. Zhang, L.; Yang, F.; Daniel Zhang, Y.; Zhu, Y.J. Road crack detection using deep convolutional neural network. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3708–3712. [CrossRef]
7. Cha, Y.J.; Choi, W.; Büyüköztürk, O. Deep learning-based crack damage detection using convolutional neural networks. *Comput.-Aided Civ. Infrastruct. Eng.* **2017**, *32*, 361–378. [CrossRef]
8. Chen, F.C.; Jahanshahi, M.R. NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion. *IEEE Trans. Ind. Electron.* **2018**, *65*, 4392–4400. [CrossRef]
9. Yang, X.; Li, H.; Yu, Y.; Luo, X.; Huang, T.; Yang, X. Automatic pixel-level crack detection and measurement using fully convolutional network. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 1090–1109. [CrossRef]
10. Huyan, J.; Li, W.; Tighe, S.; Xu, Z.; Zhai, J. CrackU-net: A novel deep convolutional neural network for pixelwise pavement crack detection. *Struct. Control Health Monit.* **2020**, *27*, e2551. [CrossRef]
11. Zou, Q.; Zhang, Z.; Li, Q.; Qi, X.; Wang, Q.; Wang, S. DeepCrack: Learning hierarchical convolutional features for crack detection. *IEEE Trans. Image Process.* **2019**, *28*, 1498–1512. [CrossRef]
12. Li, S.; Zhao, X.; Zhou, G. Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. *Comput.-Aided Civ. Infrastruct. Eng.* **2019**, *34*, 616–634. [CrossRef]
13. 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.* **2021**, *36*, 1094–1113. [CrossRef]
14. Zheng, Y.; Gao, Y.; Lu, S.; Mosalam, K.M. Multistage semisupervised active learning framework for crack identification, segmentation, and measurement of bridges. *Comput.-Aided Civ. Infrastruct. Eng.* **2022**, *37*, 1089–1108. [CrossRef]
15. Chen, J.; Lu, W.; Lou, J. Automatic concrete defect detection and reconstruction by aligning aerial images onto semantic-rich building information model. *Comput.-Aided Civ. Infrastruct. Eng.* **2023**, *38*, 1079–1098. [CrossRef]
16. Zou, Q.; Cao, Y.; Li, Q.; Mao, Q.; Wang, S. CrackTree: Automatic crack detection from pavement images. *Pattern Recognit. Lett.* **2012**, *33*, 227–238. [CrossRef]
17. Fan, L.; Li, S.; Li, Y.; Li, B.; Cao, D.; Wang, F.Y. Pavement cracks coupled with shadows: A new shadow-crack dataset and a shadow-removal-oriented crack detection approach. *IEEE/CAA J. Autom. Sin.* **2023**, *10*, 1593–1607. [CrossRef]
18. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 27, pp. 2672–2680.
19. Xu, B.; Liu, C. Pavement crack detection algorithm based on generative adversarial network and convolutional neural network under small samples. *Measurement* **2022**, *196*, 111219. [CrossRef]
20. 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. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 4474–4481. [CrossRef]

21. Branikas, E.; Murray, P.; West, G. A novel data augmentation method for improved visual crack detection using generative adversarial networks. *IEEE Access* **2023**, *11*, 22051–22059. [CrossRef]
22. Jin, T.; Ye, X.W.; Li, Z.X. Establishment and evaluation of conditional GAN-based image dataset for semantic segmentation of structural cracks. *Eng. Struct.* **2023**, *285*, 116058. [CrossRef]
23. Sauer, A.; Chitta, K.; Müller, J.; Geiger, A. Projected GANs converge faster. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 17480–17492.
24. Tran, D.; Ranganath, R.; Blei, D. Hierarchical implicit models and likelihood-free variational inference. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30, pp. 5523–5533.
25. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30, pp. 6626–6637.
26. Hsieh, Y.A.; Tsai, Y.J. Machine learning for crack detection: Review and model performance comparison. *J. Comput. Civ. Eng.* **2020**, *34*, 04020038. [CrossRef]
27. Ye, X.W.; Jin, T.; Li, Z.X.; Ma, S.Y.; Ding, Y.; Ou, Y.H. Structural crack detection from benchmark data sets using pruned fully convolutional networks. *J. Struct. Eng.* **2021**, *147*, 04721008. [CrossRef]
28. Pak, M.; Kim, S. Crack detection using fully convolutional network in wall-climbing robot. In *Proceedings of the Advances in Computer Science and Ubiquitous Computing*; Springer: Singapore, 2021; pp. 267–272. [CrossRef]
29. Shijie, J.; Ping, W.; Peiyi, J.; Siping, H. Research on data augmentation for image classification based on convolution neural networks. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 4165–4170. [CrossRef]
30. Cao, J.; Hou, L.; Yang, M.H.; He, R.; Sun, Z. ReMix: Towards image-to-image translation with limited data. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 15013–15022. [CrossRef]
31. Choi, Y.; Uh, Y.; Yoo, J.; Ha, J.W. StarGAN v2: Diverse image synthesis for multiple domains. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 8185–8194. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.