# CrackGauGAN: Semantic Layout-based Crack Image Synthesis for Automated Crack Segmentation

**Honghu Chu[1, 2] ,Weiwei Chen[1,3 *] , Lu Deng[2]**

[1] Bartlett School of Sustainable Construction, University College London, UK
[2] College of Civil Engineering, Hunan University, China
[3] Division D Civil Engineering, Engineering Department, University of Cambridge, UK
weiwei.chen@ucl.ac.uk, honghu.chu.22@ucl.ac.uk, denglu@hnu.edu.cn

**Abstract**

Automated crack inspection, particularly deep learning (DL)-based crack segmentation, is crucial for the effective and efficient maintenance, repair, and operation of civil infrastructure. However, the performance of DL-based segmentation methods is often limited by the scarcity of pixel-wise labeled crack images. This paper presents CrackGauGAN, an automated crack image synthesis network that can be used to generate realistic and diverse crack image and mask pairs, which are instrumental in improving the performance of DL-based crack segmentation models. The CrackGauGAN is developed with three customized improvements based on the original GauGAN architecture. Firstly, a Criminisi-based crack image inpainting operator is introduced before the image encoder, enabling the exclusion of crack noise interference during background color feature extraction. Secondly, a background texture extraction method is proposed, assisting the SPADE-based generator in decoupling background textures as prior information. Lastly, an adaptive pseudo-augmentation strategy is introduced in the discriminator, allowing the model to be effectively trained on small-scale crack datasets. Ablation studies are conducted to prove the effectiveness of each component, and further crack image generation experiments demonstrate that the CrackGauGAN can synthesize various cracks with excellent diversity and fidelity. The CrackGauGAN-generated crack images show average improvements of over 1.97 and 7.91 in the Inception Score (IS) and Fréchet Inception Distance (FID), respectively, compared to the previously most advanced GauGAN and Pix2PixHD. As a fully automated crack image mask pair generation architecture, the CrackGauGAN can be used to provide reliable data support for the application of DL-based segmentation models in crack inspection tasks.

**Keywords –**

**Deep Learning; Generative Adversarial Network; Crack Image Synthesis; Feature Decoupling; Crack Segmentation**

## 1 INTRODUCTION

In recent decades, the aging of bridges and escalating traffic loads have markedly intensified the issue of structural safety [1, 2]. Cracks, as one of the most prevalent and evident indicators of structural safety [3], are particularly noteworthy. Early detection and prompt maintenance of these cracks can substantially lower maintenance expenses over the bridge's operational lifespan.

Traditionally, crack detection has been conducted through visual inspections by qualified experts, a process that is both costly and time-consuming [4, 5]. This approach struggles to keep pace with the escalating global demand for bridge inspections, particularly for long-span structures [6]. However, the advent of visual sensors in recent decades has facilitated the continuous acquisition of image data from civil infrastructure, utilizing automated robots, unmanned aerial vehicles (UAVs), camera-equipped vehicles, and fixed surveillance cameras on bridges [7]. The emergence of visual recognition models based on deep learning (DL) technologies offers the potential for accurate and efficient identification of cracks in these images, garnering significant interest from both industry and academia [8].

Recent research in bridge crack image processing has leveraged DL techniques, achieving significant advancements in crack classification, object detection, and segmentation. These techniques are increasingly recognized as the most promising solution for automating and streamlining detection tasks, potentially replacing manual labor [9-11]. While DL technologies represent the state-of-the-art in the field of pixel-level crack detection, their efficacy hinges on the availability of extensive source data for accurate network training. Limited datasets can lead to network overfitting, where the model excels on training data but exhibits markedly reduced performance in unfamiliar environments [12].

To mitigate the issue of limited training data, researchers have explored data augmentation techniques to expand crack datasets [13, 14]. These methods fall into two categories: basic image processing and DL-based approaches [15, 16]. Basic image processing techniques, such as flipping, cropping, and rotating, generate new crack images that retain the original's semantic information [17]. However, these methods do not significantly enhance the diversity of crack samples, thus offering limited improvement in the accuracy of crack image recognition. In contrast, DL-based data augmentation algorithms, particularly Generative Adversarial Networks (GANs), can generate diverse, high-resolution images that mirror the distribution of the original dataset, thereby effectively enriching the variety of training samples [12].

Exploring GANs for synthesizing crack images with targeted domains has always been a topic worthy of discussion [18]. Numerous studies have shown that adding GAN-generated images to the original training data can make DL-based recognition models more accurate [19]. It is important to note that the generated images can be used to improve the performance of DL-based segmentation models on the condition that the images have detailed pixel-level annotations. However, carrying out the annotation process is an extremely tedious task. Therefore, to significantly enhance the convenience of using generated images, this study aims to use the GAN to generate realistic crack images while automatically obtaining their corresponding masks. To the best of the authors' knowledge, similar studies have not been reported. The main contributions to this work are as follows:

This study proposes a novel Generative Adversarial Network (GAN) architecture, named CrackGauGAN, which facilitates the generation of realistic crack images solely from semantic maps. The core innovation lies in integrating crack inpainting and texture priors into the GauGAN architecture [20], enabling the network to pre-decouple three key features in crack image generation - crack morphology, texture, and color. This innovation allows the SPADE (Spatially-Adaptive Normalization) blocks in GauGAN, originally designed for natural scene image generation, to be used for creating clear and realistic crack images. Furthermore, given the extensive image data required for parameter training due to the large number of parameters in the original GauGAN architecture, the authors implemented an adaptive data augmentation strategy in the discriminator. This approach allows for effective training of the CrackGauGAN on a limited dataset. The well-trained model is capable of automatically generating crack image and mask pairs, which can be directly used to enhance the performance of DL-based segmentation models, thereby promoting the application of DL-based crack segmentation methods in practical engineering.

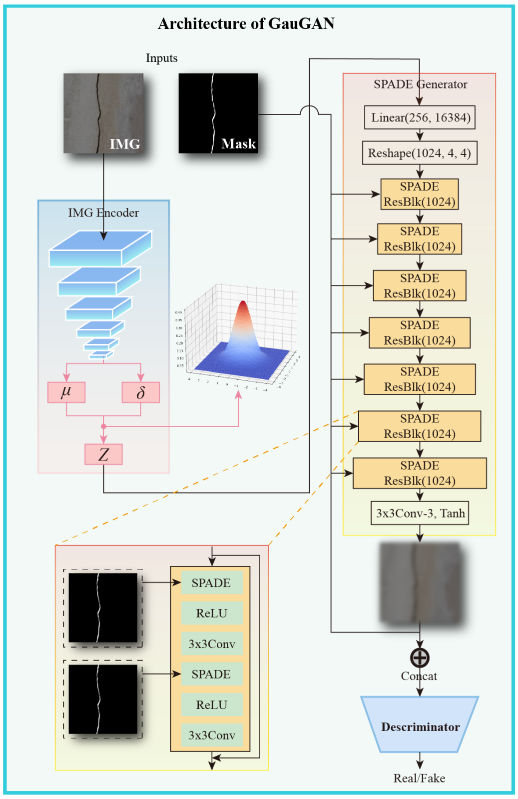


Figure 1. The overview architecture of the original GauGAN

## 2 METHODOLOGY

### 2.1 Overview of GauGAN

The CrackGauGAN model is built upon the GauGAN model [20], and it is necessary to explain the composition of the GauGAN model when introducing Crack-GauGAN. The GauGAN architecture, as shown in Figure 1, mainly consists of three key components: Image Encoder, SPADE-based Generator, and Discriminator.

The Image Encoder is designed to extract the mean $\mu$ and variance $\delta$ related to the color feature distribution from the real image. Then, the extracted mean $\mu$, variance $\delta$, and the the Gaussian distribution $x$ would be denormalized, ultimately obtaining a random vector $z$ that contains the color information of the real image.

The function of the SPADE-based Generator is to receive the random vector $z$ generated in the previous step and enhance the realism of the pixels in the generated image by continuously using the semantic map.

The Discriminator is customized to process the tensor resulting from the integration of the semantic map and the generated image. It executes conditional discrimination across multiple scales, enabling the effective assessment of both global features, like background color and crack distribution, and local features, including texture details, in the generated image. This functionality is crucial for ensuring the image's overall clarity.

Through continuous training involving a contest between the generator and discriminator, GauGAN ultimately produces realistic images that match the target distribution locations in the input semantic maps.

## 2.2 Revised Architecture for Crack Images

Unlike natural scene images with fixed shapes and significant semantic arrangement relationships, crack images consist only of cracks with random morphological distributions and backgrounds lacking semantic information. Consequently, traditional GauGAN, when dealing with such hard samples, tends to produce blurred and artifact-ridden images due to the difficulty in adequately decoupling the deep semantic features of cracks and backgrounds during the training phase. Additionally, the limited pixel-level annotated open-sourced crack datasets make GauGAN prone to overfitting during training. The authors address these issues effectively through three customized designs, and the revised architecture is shown in Figure 2, where the customized components are highlighted in red in the revised architecture.

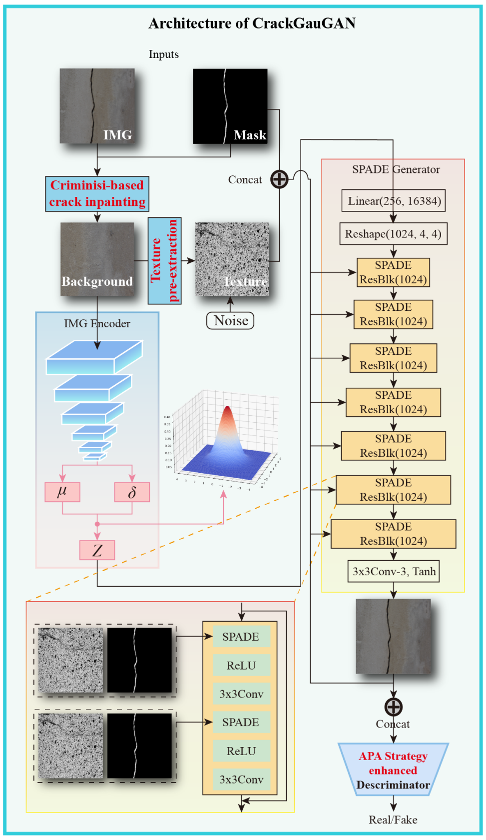


Figure 2. The overview architecture of the proposed CrackGauGAN

### 2.2.1 Revision 1: Criminisi-based crack inpainting

Considering that both cracks and backgrounds lack distinct semantic information that can be effectively differentiated by the image encoder, the random vector $z$ often fails to accurately represent the color information of the image target due to feature coupling. To address this, the authors introduce a crack inpainting operation before the original image is input into the image encoder. This operation repairs the crack areas in the background image, thereby eliminating the interference of crack pixels in the extraction of the background color information vector.

To effectively repair the crack areas, it is necessary to ensure that the background texture of the repaired crack areas can transition smoothly. In this study, the Criminisi method [21] is employed. The Criminisi method, with its advantages in diffusion-based repair and texture synthesis, performs well on images with large missing areas and those composed of textures and structures. The specific implementation method is as follows:

1. Define the repair area: First, identify the crack areas in the crack image that need to be repaired from the input semantic labels;
2. Initialize priority: The algorithm then assigns priorities to the edge pixels of the area to be repaired based on structural information like pixel gradients and texture information based on the proportion of known pixels;
3. Select the source area: The algorithm searches for an undamaged area in the image that best matches the texture and structure at the current highest priority edge as the source area for repair;
4. Texture and structure replication: Copy the selected source area to the edge location with the highest priority, thereby filling a part of the repair area;
5. Update priorities and repair area: After each fill, the algorithm updates the edges of the repair area and the priorities of the corresponding pixels, and repeats steps 3 to 5 until the entire repair area is filled.

### 2.2.2 Revision 2: DTCWT and PSO-based background texture pre-extraction

The visual quality judgment of crack images by the naked eye primarily relies on three indicators: background color, crack distribution, and background texture. The background color and crack distribution have already been individually represented by the image encoder and the semantic map, respectively. To enable the network to effectively control all three critical indicators that determine the generation of crack images, it is necessary to add texture prior information to the network to help decouple texture information in advance.

Specifically, this study introduces an image texture feature extraction algorithm based on Dual-Tree Complex Wavelet Transform (DTCWT) and Particle

Swarm Optimization (PSO), designed to introduce texture prior information of the crack background into each SPADE block. Compared to traditional image texture feature extraction algorithms like Canny and Sobel, the texture information extraction method proposed in this study effectively leverages the efficient texture analysis capability of DTCWT and the global search advantage of PSO, thereby enhancing the accuracy and efficiency of image texture feature extraction. The implementation details of the DTCWT and PSO methods can be found in [22].

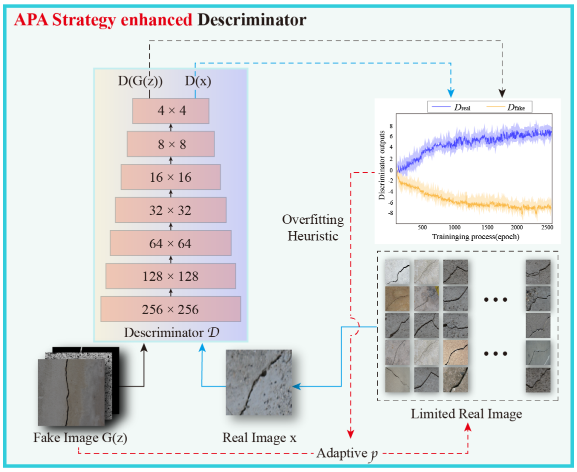


Figure 3. Workflow for adaptive pseudo-augmentation strategy in the discriminator

### 2.2.3 Revision 3: Adaptive pseudo-augmentation strategy in the Discriminator

To mitigate the potential overfitting issue due to the insufficiency of initial training source crack images, a new data augmentation strategy called adaptive pseudo-augmentation strategy (APA) is introduced in the discriminator, as shown in Figure 3.

Researchers have been challenged by obtaining well-trained generative models based on carefully constructed GAN frameworks, with traditional methods relying on a large number of training images to ensure the model avoids overfitting. For the GauGAN architecture, designed for generating natural scene images, open-source datasets like Pascal VOC 2012, COCO, and Cityscapes, which contain tens of thousands of pixel-level annotated images, have alleviated the data issue to some extent. However, due to the difficulties in collecting crack images and the reliance on professional personnel for annotation, there is no such large-scale open-source dataset available for cracks. Therefore, effectively training CrackGauGAN on small sample datasets becomes the problem to be addressed in this section. To this end, the authors designed a data augmentation pipeline for the limited crack training samples, called the adaptive pseudo-augmentation Strategy, first proposed by Huang et al. [23]. It can dynamically adjust the intensity of augmentation based on the degree of overfitting in the field of medical imaging generation, without leaking augmentation patterns. Its effectiveness has been proven through the ablation study described in section 4.1.3.

# 3 IMPLEMENTATION DETAILS AND EXPERIMENTS

## 3.1 Datasets

The open-source crack segmentation dataset HRCD-282, previously established by the authors, was used as the data source for model training and evaluation. From HRCD-282, 1200 crack patch images with a resolution of 256×256 and their corresponding crack labels were cropped from the included HR crack images to serve as training data for CrackGauGAN. Additionally, to assess the quality of images generated by the model, 1200 non-crack patch images with the same resolution of 256×256 were collected from the backgrounds of the selected original HR crack images. These 1200 non-crack background images, along with crack labels of the same size, were input into the well-trained CrackGauGAN model to generate pseudo-crack images under the corresponding backgrounds.

## 3.2 Implementation Details

In this study, all experiments were conducted on a high-performance workstation equipped with an Intel Core i9-9820X CPU and NVIDIA RTX 3090 Ti GPU. The workstation runs on the Ubuntu 20.04 LTS operating system, providing a stable and efficient computing environment. Furthermore, to ensure the fairness of comparative experiments, all networks involved in the comparison were implemented under the PyTorch framework.

To optimize the training effectiveness of the Crack-GauGAN model, the authors carefully selected the following key hyperparameters. The initial learning rate was set at 0.0001 to ensure stable gradient descent, and a learning rate decay strategy was adopted to address potential overfitting during training. Considering the limitations of hardware resources, the batch size was cautiously set to 8 to balance memory consumption and training efficiency. Additionally, the Adam optimizer was chosen, with its beta1 and beta2 parameters set to 0.5 and 0.999, respectively, to take advantage of its adaptive learning rate. The weight decay parameter was set at $1\times10^{-4}$ to further prevent model overfitting.

In terms of the update strategy for the generator and discriminator, a 1:1 ratio was followed to ensure balanced optimization of both during training. The weighting of the loss function was also carefully adjusted to balance the impact of adversarial loss, feature matching loss, and VGG perceptual loss, thereby optimizing the overall performance of the model. Considering the importance of crack image details and the limitations of computational resources, a training image size of 256 × 256 was chosen. Finally, to ensure that the model learned sufficiently and converged, the number of training epochs was set to 900. The selection of these hyperparameters was based on a comprehensive analysis of previous studies and the results of preliminary experiments, aiming to achieve the best training effect and image quality.

### 3.3 Evaluation Indicators

To comprehensively evaluate the model performance of CrackGauGAN, the authors conducted qualitative assessments through visualized generation results of all models and also introduced two quantitative evaluation metrics: Inception Score (IS) and Fréchet Inception Distance (FID). IS evaluates the diversity and clarity of the generated images, while FID measures the distance between the generated images and real images in the feature space.

## 4 EXPERIMENTS AND RESULTS ANALYSIS

### 4.1 Ablation Studies

#### 4.1.1 Ablation study of the Criminisi-based crack inpainting

Table 1 quantitatively shows the ablation study for the image background inpainting. It can be found that by removing the background inpainting, the training process becomes more difficult to converge, resulting in a larger FID score and a smaller IS score, which indicates a decrease in the realism and clarity of the generated images. This is because crack images differ from natural scene images in that there is no significant color contrast difference between the foreground and background, and neither possesses a fixed topological structure. This makes it challenging for the image encoding architecture to effectively decouple their deep semantic information. As a result, the network's process of extracting background color features is easily disrupted by crack features, leading to a significant decline in the quality of color feature extraction. Experimental results confirm that repairing crack pixels in the image background effectively mitigates this issue.

#### 4.1.2 Ablation study of the DTCWT and PSO-based texture pre-extraction

This ablation study primarily focuses on the impact of the texture prior information intensity, extracted by the proposed DTCWT and PSO operations, on the quality of image generation. Figure 4 illustrates these comparisons across texture information intensities of 0%, 25%, 50%, 75%, and 100%. A visual evaluation of the reconstruction results in Figure 4 revealed that the model utilizing 100% texture information extraction intensity achieved the closest resemblance to the original image. Furthermore, it was observed that at texture information extraction intensities below 50%, the GAN struggled to independently decouple crack features from background texture information. This challenge is attributed to the excessive redundancy in low semantic information, a consequence of the random data distribution in crack images. Consequently, early decoupling of high-intensity texture information significantly enhances the network's proficiency in distinguishing and capturing both background semantic information and crack features, a critical factor for reconstructing high-quality crack images.

Table 1. The impact of the proposed inpainting operation on the quality of generated images and the difficulty of the model convergence

| Criminisi-based crack inpainting opeartion | IS(↑) | FID(↓) | Iterations (epoch) |
|---|---|---|---|
| W/O | 3.43 | 34.56 | 1700 |
| W/ | 5.26 | 26.33 | 900 |

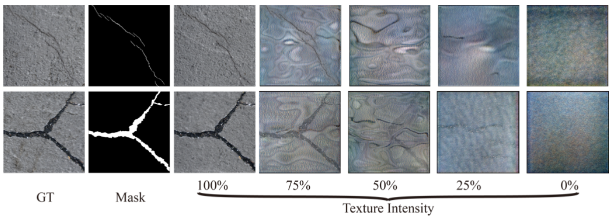


Figure 4. Visualization results of crack images generated by models with added texture prior information of varying intensities

#### 4.1.3 Ablation study of adaptive pseudo-augmentation Strategy

Figure 5 summarizes the impact of the adaptive pseudo-augmentation (APA) strategy on both the generator and discriminator losses in the CrackGauGAN. The implementation of the APA strategy led to a notable reduction in the oscillation amplitudes of both losses during training on a small-scale dataset. This indicates that the APA strategy effectively stabilized the feature learning of both the discriminator and generator within the bounds of maximum gradient convergence. This is particularly significant for the discriminator loss, where the extremely limited training data can rapidly lead to local overfitting, thereby creating a false impression of model convergence. The APA strategy successfully counters this issue. Analysis of the discriminator (D) loss and generator (G) loss behaviors suggests that the APA strategy not only prevents overfitting in the discriminator but also mitigates the problem of gradient vanishing, thereby ensuring the generator's continuous learning.

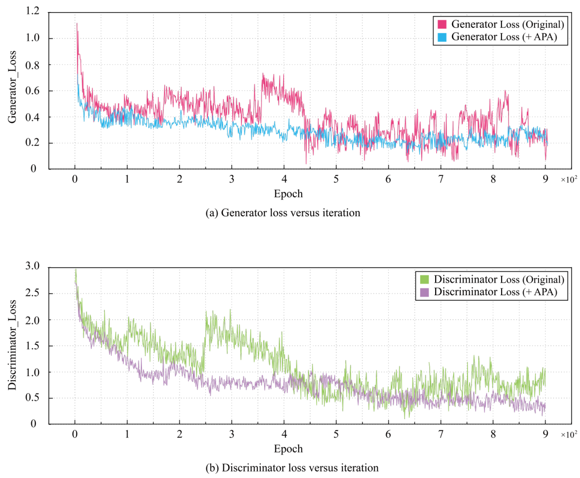


Figure 5. The behavior of D Loss and G Loss of the proposed architecture with and without APA strategy in the training phase

## 4.2 Performance Comparison with Current State-of-the-art Methods

To demonstrate the advancement of the method proposed in this study, it was compared with the current state-of-the-art semantic image synthesis methods: GauGAN [23] and pix2pixHD model [24]. The Ground Truth (GT) used for evaluating the quality of the generated images originates from the HR crack images collected by the authors, as described in section 3.1. For each GT, the corresponding crack mask and the non-crack background images were used as inputs for the models. Additionally, it is important to note that to ensure fairness in comparison, all models involved in the experiment were trained using the default hyperparameters provided by their original authors.

**Qualitative Results Analysis:** Figure 6 illustrates some of the crack images generated by all the methods involved in the comparison. It is intuitively evident from the figure that all the compared methods can generate crack images broadly consistent with the distribution of the crack semantic map. Among these, the crack images synthesized through CrackGauGAN are closer to real crack images. Specifically, they exhibit fewer artifacts, richer details, and clearer edges in terms of visual quality. As for the Pix2PixHD, due to the lack of decoupling operations for cracks and background textures during the training phase, tends to lose background texture information in the generated images. Although GauGAN, which employs SPADE, can utilize the spatially adaptive mechanism to improve the entangled coupling between background and crack features, the limited crack training samples restrict its performance, resulting in the generated crack background textures being blurred. Further observation of the crack images generated by the CrackGauGAN reveals that different levels of noise input do not affect the clarity of the images or the distribution of cracks in the images, but only alter the distribution of background textures. This is significant for enhancing the diversity of the images.

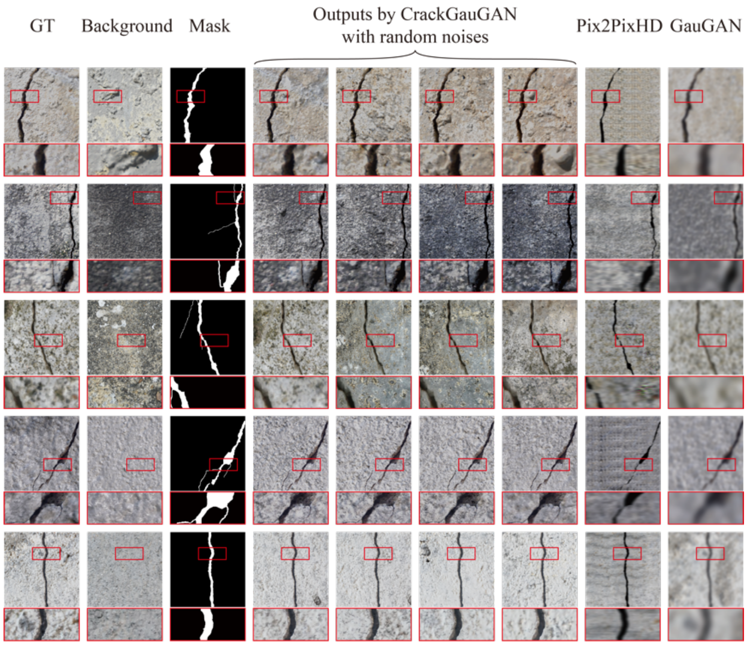


Figure 6. Qualitative comparison of CrackGauGAN with two state-of-the-art image synthesis methods

Table 2. Quantitative results of the quality of images generated by CrackGauGAN with different intensity noise added and the current state-of-the-art methods

| Model | The intensity of random noise added | IS | FID |
|---|---|---|---|
| Pix2PixHD | / | 4.20 | 31.65 |
| GauGAN | / | 4.78 | 29.38 |
| CrackGauGAN | 20% | 6.47 | 22.74 |
| | 40% | 6.53 | 22.09 |
| | 60% | 6.43 | 22.76 |
| | 80% | 6.39 | 23.12 |

**Quantitative Results Analysis:** Table 2 reports the quantitative results of the crack image quality generated using different methods. As can be seen from Table 2, adding different proportions of random noise has almost no impact on the evaluation metrics. This can be inferred from the qualitative analysis results shown in Figure 4, as the random noise only affects the distribution of the background texture and does not impact the clarity of the background texture or other details related to image quality. Moreover, under four types of noise inputs, the IS and FID of the images generated by GauGAN fluctuate within the range of 6.46±0.07 and 22.61±0.52, respectively. This represents an average improvement of over 1.97 and 7.91 compared to the images generated by Pix2PixHD and GauGAN, respectively. This effectively confirms the advanced nature of the method proposed in this study.

# 5 CONCLUSION

This paper presents a novel CrackGauGAN, a semantically-driven generative adversarial network specifically tailored for crack image generation. By effectively decoupling three key features for the generator, including background color, background texture, and crack morphology, the CrackGauGAN is capable of generating realistic crack images with high fidelity and diverse data distribution types. Its performance surpasses that of the most advanced semantic layout-based GANs, with average improvements in Inception Score (IS) and Fréchet Inception Distance (FID) exceeding 1.97 and 7.91, respectively, compared to GauGAN and Pix2PixHD.

The proposed generative model can quickly generate a large number of realistic crack image datasets with pixel-level labels for in-service bridges, overcoming the challenge of insufficient training samples with similar data distribution types. In the future, this approach will be extended to create customized crack image datasets for additional bridges and train corresponding crack detection models. Such models will facilitate accurate and efficient intelligent detection methods in engineering practice, including bridges, hydropower projects and historical buildings, enabling evidence-based infrastructure maintenance and management decisions.


# Acknowledgement

This work is supported by Horizon Europe project D-HYDROFLEX (Project ID: 101122357) and Horizon Europe project INHERIT (Project ID: 101123326).


# References


[1] Zhang C, Chang C-C, Jamshidi M. Concrete Bridge Surface Damage Detection Using A Single-Stage Detector. *Computer-Aided Civil And Infrastructure Engineering*, 35(4):389-409, 2020.

[2] Abu Dabous S, Feroz S. Condition Monitoring Of Bridges With Non-Contact Testing Technologies. *Automation In Construction*, 116:103224, 2020.

[3] Yu Z, Shen Y, Shen C. A Real-Time Detection Approach For Bridge Cracks Based On Yolov4-FPM. *Automation In Construction*, 122:103514, 2021.

[4] Jiang S, Zhang J. Real-Time Crack Assessment Using Deep Neural Networks With Wall-Climbing Unmanned Aerial System. *Computer-Aided Civil And Infrastructure Engineering*, 35(6):549-64, 2020.

[5] Rodriguez-Lozano FJ, LeÓn-GarcÍa F, GÁmez-Granados JC, Palomares JM, Olivares J. Benefits Of Ensemble Models In Road Pavement Cracking Classification. *Computer-Aided Civil And Infrastructure Engineering*, 35(11):1194-208, 2020.

[6] Guo F, Qian Y, Liu J, Yu H. Pavement Crack Detection Based On Transformer Network. *Automation In Construction*, 145:104646, 2023.

[7] Bianchi E, Abbott Amos L, Tokekar P, Hebdon M. COCO-Bridge: Structural Detail Data Set For Bridge Inspections. *Journal Of Computing In Civil Engineering*, 35(3):04021003, 2021.

[8] Alipour M, Harris Devin K, Miller Gregory R. Robust Pixel-Level Crack Detection Using Deep Fully Convolutional Neural Networks. *Journal Of Computing In Civil Engineering*, 33(6):04019040, 2019.

[9] Corbally R, Malekjafarian A. A Deep-Learning Framework For Classifying The Type, Location, And Severity Of Bridge Damage Using Drive-By Measurements. *Computer-Aided Civil And Infrastructure Engineering*, 37(3):1163-1188, 2023.

[10] Yang Q, Shi W, Chen J, Lin W. Deep Convolution Neural Network-Based Transfer Learning Method For Civil Infrastructure Crack Detection. *Automation In Construction*. 116:103199, 2020.

[11] Zhou S, Canchila C, Song W. Deep Learning-Based Crack Segmentation For Civil Infrastructure: Data Types, Architectures, And Benchmarked Performance. *Automation In Construction*, 146:104678, 2023.

[12] Shi J, Liu W, Zhou G, Zhou Y. Autoinfo GAN: Toward A Better Image Synthesis GAN Framework For High-Fidelity Few-Shot Datasets Via NAS And Contrastive Learning. *Knowledge-Based Systems*, 276:110757, 2023.

[13] Liu B, Zhang T, Yu Y, Miao L. A Data Generation Method With Dual Discriminators And Regularization For Surface Defect Detection Under Limited Data. *Computers In Industry*, 151:103963, 2023.

[14] Liu Q, Zhang Q, Liu W, Chen W, Liu X, Wang X. WSDS-GAN: A Weak-Strong Dual Supervised Learning Method For Underwater Image Enhancement. *Pattern Recognition*, 143:109774, 2023.

[15] Xu G, Yue Q, Liu X. Real-Time Monitoring Of Concrete Crack Based On Deep Learning Algorithms And Image Processing Techniques. *Advanced Engineering Informatics*, 58:102214, 2023.

[16] Li S, Zhao X. High-Resolution Concrete Damage Image Synthesis Using Conditional Generative Adversarial Network. *Automation In Construction*, 147:104739, 2023.

[17] Liu X, Pedersen M, Wang R. Survey Of Natural Image Enhancement Techniques: Classification, Evaluation, Challenges, And Perspectives. *Digital Signal Processing*, 127:103547, 2022.

[18] Zhang G, Cui K, Hung T-Y, Lu S, Editors. Defect-GAN: High-Fidelity Defect Synthesis For Automated Defect Inspection. *Proceedings Of The IEEE/CVF Winter Conference On Applications Of Computer Vision*, Pages 35–43, Virtual Online, 2021.

[19] Jin T, Ye X, Li Z. Establishment And Evaluation Of Conditional GAN-Based Image Dataset For Semantic Segmentation Of Structural Cracks. *Engineering Structures*, 285:116058, 2023.

[20] Park T, Liu M-Y, Wang T-C, Zhu J-Y, Editors. Semantic Image Synthesis With Spatially-Adaptive Normalization. *Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition*, Pages 56–64, Long Beach, America, 2019.

[21] Yao F. Damaged Region Filling By Improved Criminisi Image Inpainting Algorithm For Thangka. *Cluster Computing*, 22:13683-91, 2019.

[22] Chi J, Eramian M. Enhancing Textural Differences Using Wavelet-based Texture Characteristics Morphological Component Analysis: A Preprocessing Method for Improving Image Segmentation. *Computer Vision and Image Understanding*, 158:49-61, 2017.

[23] Huang X, Belongie S, editors. Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization. *Proceedings of the IEEE International Conference on Computer Vision*, page: 77-85, Venice, Italy, 2017.

[24] Wang T-C, Liu M-Y, Zhu J-Y, Tao A, Kautz J, Catanzaro B, editors. High-resolution Image Synthesis and Semantic Manipulation with Conditional Gans. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, page: 121-129, Salt Lake City, America, 2018.