Research article

A pavement crack synthesis method based on conditional generative adversarial networks

Hui Yao[1,*], Yuhan Wu[1], Shuo Liu[1], Yanhao Liu[1] and Hua Xie[2]

[1] Beijing Key Laboratory of Traffic Engineering, College of Metropolitan Transportation, Faculty of Architecture, Civil and Transportation Engineering, Beijing University of Technology, Beijing 100124, China

[2] Yan'an University, Yan'an 716000, China

* Correspondence: Email: huiyao@mtu.edu.

Abstract: A Generative Adversarial Network (GAN) based asphalt pavement crack image generation method was proposed to improve the dataset size of the road images. Five open-source road crack datasets were leveraged to construct an image dataset, which contained two labels - transverse cracks and longitudinal cracks. The constructed dataset was used to facilitate crack detection and classification research by providing a diverse collection of labeled crack images derived from multiple public sources. The network structure of fully connected, convolutional and attention mechanisms based on the Conditional Generative Adversarial Network (CGAN) was used in this project. The purpose of this study was to train a generative model on selected categories of input pavement crack images and generate realistic crack images of those categories. We aim to tune the parameters of the GAN and optimize hyperparameters to improve the realism possibility of generated images. It also explored the generated images with different sizes and evaluated the performance of networks with different architectures. In particular, we analyzed the structural characteristics of conditional GAN. Results demonstrated that the Self-Attention Generative Adversarial Networks (SAGAN) model, which combines self-attention mechanisms with CGAN, can effectively address challenges related to limited crack image data and the inability to selectively generate images from specific categories. By conditioning the generator on category information, the SAGAN model was able to generate high-quality images while focusing on the target categories. Overall, the self-attention and conditional aspects of the SAGAN framework helped improve the generation of realistic pavement crack images.

1. Introduction

In recent years, pavement maintenance issues have received more and more attention. With the increasing mileage of global maintenance [1], and the amazingly rapid development of artificial intelligence, especially deep learning-based methods, automatic pavement distress detection methods [2–6], and pavement distress segmentation methods [7], have been continuously proposed to address the challenges of inspecting vast transportation infrastructure in a timely and cost-effective manner. As one of the most common pavement distresses, cracks pose risks to the structural integrity of the pavement surface if not addressed promptly through appropriate repair measures. Left unrepaired, cracks have the potential to develop into more severe issues that can further impact the service life and level of the pavement. An adequate amount of data is a prerequisite for deep learning to be feasible. Although deep learning has some advantages in the field of image vision, it relies on a massive number of samples to achieve high-precision training results. The size of the dataset must be guaranteed to ensure robustness and generalization. Compared with the modification of the model structure and optimization of the parameters, the augmentation of the data volume is usually the most expeditious avenue for enhancing performance.

In addition to acquiring more real-world pavement images through field collection, the Generative Adversarial Network (GAN) was proposed by Goodfellow et al. [8]. The feature distribution of the samples was learned and the learned features were combined to generate realistic images. Generative Adversarial Networks (GANs) are a type of deep learning model that comprises two components-a Generator and a Discriminator--which interact in a game-like manner. The objective of the Generator is to create data that closely resembles authentic examples that can deceive the Discriminator. On the other hand, the Discriminator aims to discern as accurately as possible between the real data and the synthetic data developed by the Generator. The generator and discriminator engage in a Minimax Game, where the former aims to maximize the probability of the latter misclassifying the generated samples, while the latter aims to minimize the said probability. During GAN training, certain issues may arise, such as Mode Collapse and unstable training, which require the implementation of various techniques and improved GAN structures for resolution. Since then, the GAN model has been continuously improved and many variants have been generated based on the original network structure, but there are many problems in the algorithms, such as difficulty in training the original GAN, the inability of the loss function to guide the model training process and poor diversity of the generated samples. The GAN model was improved by Arjovsky et al. [9], targeting the original JS scattering objective function through Wasserstein distance instead and casting certain improvements in the stability of training and image quality. The progressive growing GAN (PGGAN) from Karaas et al. [10] is a training method that starts with low-resolution images and continues to complicate the model by adding new network layers to gradually learn detailed features, thus achieving accelerated and stable training. Least Squares Generative Adversarial Networks (LSGANs) were generated by Mao et al. [11] to improve the quality and stability of generated images through the least squares loss function. A style-based generative adversarial network (StyleGAN) architecture was used by Karras et al. [12] to improve the quality of generated images by providing fine-grained control over the style of generated images. The application

of GAN in road engineering is becoming increasingly prevalent. A hybrid generative adversarial network and variational autoencoder approach was employed by Pei et al. [13] to enhance the Deep Convolution GAN(DCGAN) model. Through iterative training and Adam optimization over multiple rounds, a vast number of virtual images nearly indistinguishable from real road crack photographs are obtained. A generative adversarial network (GAN) approach was also devised by Xu et al. [14] that leverages a small sample dataset of road cracks captured by unmanned aerial vehicles (UAVs) to train the GAN model for dataset expansion, generating additional synthetic images to augment the original training set. Mazzini [15] et al. used GAN to generate a semantic layout, and then a CNN-based texture synthesizer generated a new semantic image based on the generated semantic layout and obtained a reference real pavement disease image from the training set.

The application of deep learning for detecting defects in pavement images has been widely studied [2–7]. However, detection accuracy hinges on the quantity and quality of the dataset, typically in the tens of thousands, posing a critical prerequisite. Therefore, it is less commonly applied in practical engineering applications. The GAN network can effectively overcome this limitation. Although numerous scholars have made advancements in GANs in recent years, there are inherent limitations such as the inability to regulate the specific diseases generated and the uneven distribution of disease types within the image dataset, leading to reduced detection accuracy. As a result, the challenge of establishing high-quality training datasets quickly and cost-effectively has become a top priority. We utilize conditional GANs on a limited set of road crack data to generate road crack images and proposes a potential solution for researchers grappling with the dearth of diverse training samples given the use of deep learning in road crack and disease detection.

2. Exploring and categorizing GANs: Parsing principles, evolution and practice

2.1. Basic principles of GANs

Generative Adversarial Networks (GANs), one of the most inventive deep learning models in modern years, have achieved tremendous achievements in the field of computer vision. By applying game theory, it generates high-quality samples in generators and discriminators. GANs are a type of deep learning model that comprises two components - a Generator and a Discriminator - which interact in a game-like manner. The objective of the Generator is to create data that closely resembles authentic examples that can deceive the Discriminator. On the other hand, the Discriminator aims to discern as accurately as possible between the real data and the synthetic data developed by the Generator. The generator and discriminator engage in a Minimax Game, where the former aims to maximize the probability of the latter misclassifying the generated samples, while the latter aims to minimize the mentioned probability. Both keep learning in the game and ideally end up generating samples that the discriminator can't distinguish between true and false. The training process is shown in Figure 1 below.
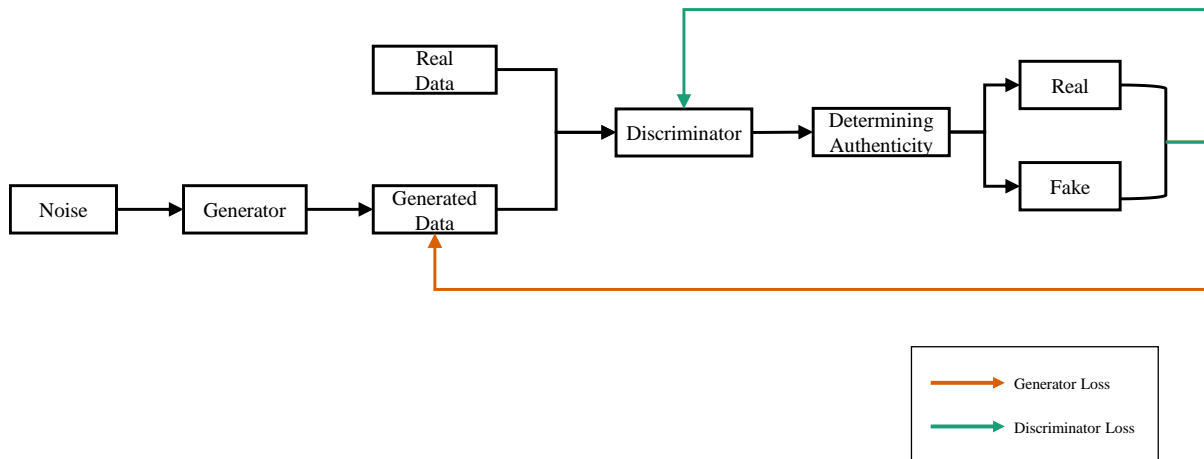
Figure 1. GAN model frame diagram.

2.1.1. The training process and objective function

In GANs, either a fixed generator is utilized to optimize the discriminator or a fixed discriminator is employed to optimize the generator, as explained in [8]. The complete formula for the GAN model is presented below.

E is the expected value of the distribution function, $E_{x \sim p_r(x)}$ is the distribution of the real sample, $E_{z-P_z(z)}$ is the distribution of the generated samples.

$$\min_{G}\max_{D}V(D,G) = E_{x \sim p_r(x)}\big[\log(D(x))\big] + E_{z-P_z(z)}\left[\log\left(1 - D\big(G(z)\big)\right)\right] \tag{1}$$

### 2.2. *Evolution of GAN in road engineering*

Since GAN was proposed by Goodfellow in 2014, it has become one of the most popular research areas in artificial intelligence and machine learning. Until now, there has been much new research on GAN in the field of road disease detection, such as WGAN and DCGAN. In this subsection, we present representative variants of GAN.

2.2.1. Deep Convolutional Generative Adversarial Networks (DCGAN)

DCGAN applies a deep convolutional neural network (CNN) as the core structure of its generator and discriminator, while the original GAN mainly employs fully connected layers. This deep convolutional structure makes DCGAN both more efficient and effective in processing image data, and the DCGAN network structure is shown in Figure 2.
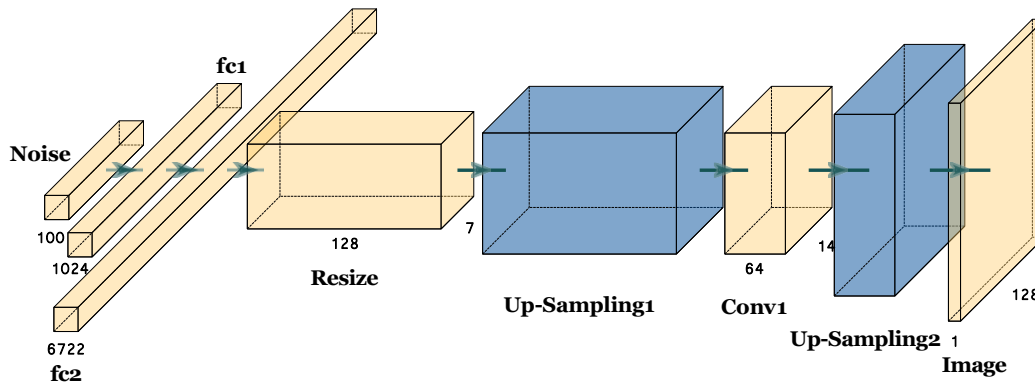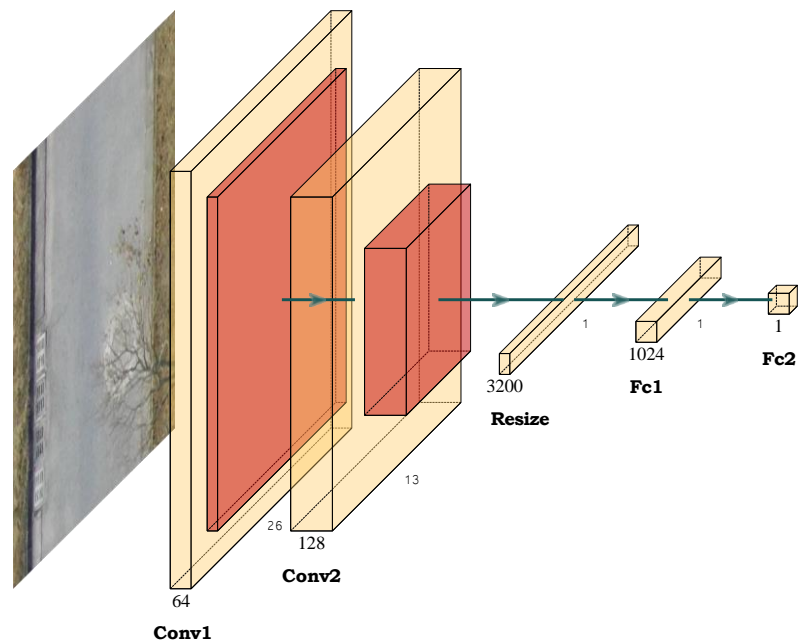
Figure 2. Schematic diagram of unconditional deep convolutional generative adversarial network. (a) Generator structure schematic diagram. (b) Discriminator structure schematic diagram.

Pei et al. [16] generated pavement crack images virtually by improving the Deep Convolutional Generative Adversarial Network (DCGAN) to overcome the lack of a sufficient number of samples in intelligent pavement inspection. A pavement crack detection model based on the Faster R-CNN network was built and trained with the original small dataset and the DCGAN-generated dataset. The experimental results show that the detection model trained with the Faster R-CNN model using the dataset expanded with images generated by DCGAN achieves an average accuracy of 90.32%, which is higher than that obtained by the traditional method evaluated using the same test dataset.

2.2.2.   WGAN [9] and WGAN-GP [17]

The original GAN minimizes the generator's loss under the near-optimal discriminator is equivalent to minimizing the JS scatter between the probability distribution of the real data and the probability distribution of the data generated by the generator, and there is a serious problem with the JS scatter: When the two distributions do not overlap, the JS scatter is zero, whereas at the beginning of the training period, the two distributions are essentially non-overlapping. Thus, if the discriminator is trained too strongly, the loss often converges to no gradient. To overcome the above issues, WGAN proposes the Wasserstein distance:

$$W(p,q) = \inf_{\gamma \in \Pi(p,q)} E_{(x,y) \sim \gamma}[||x - y||] \tag{2}$$

Equation (2) represents the Wasserstein distance between two probability distributions p and q. Where inf denotes the low bound in all possible joint distributions γ, $\Pi(p,q)$ denotes all possible joint distributions of p and q, and $||x - y||$ is the d power of the distance between two points in Euclidean space.

The advantage of utilizing the Wasserstein distance over the JS scatter is that the JS scatter is sudden and lacks gradient where there is no overlap. However, the Wasserstein distance offers a significant gradient based on distance, even in cases where the two distributions do not overlap. This regular term is the only difference between WGAN-GP and WGAN. WGAN-GP is similar to WGAN with the addition of a regular term, GP (gradient penalty), which enforces a gradient constraint.

Xu et al. [17] employed the WGAN-GP network (The network structure is shown in Figure 3 below) to gather B-Scan image data using Ground Penetrating Radar (GPR) and generate high-quality B-Scan images via unsupervised data enhancement. Subsequently, a ResNet50 [18] model was utilized to categorize subsurface basement diseases. The results of the experiments demonstrate that the data enhancement method of WGAN-GP yielded significant improvement in the classification accuracy, with the test accuracy rising to 90.85%. The overall recall rate was 90.79% and the F1 score was 72.58%. The high-quality GPR images, produced by the combination of traditional data enhancement and WGAN-GP, have significantly bettered the classification accuracy or underground stress detection in an efficient manner and also addressed the problem of category imbalance.
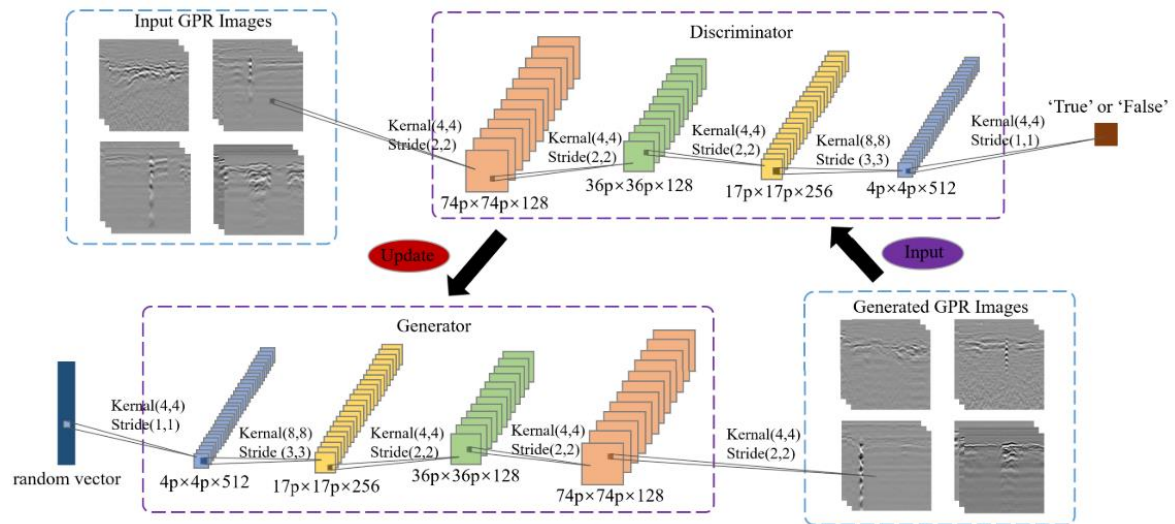
Figure 3. Architecture and training process of WGAN-GP model [19].

2.2.3. Style Based Generative Adversarial Networks (StyleGAN) [12]

In recent years, Generative Adversarial Networks (GANs) have demonstrated significant improvements in image resolution and quality. However, many research efforts perceive generators as a black box and there is limited knowledge regarding the image generation process, controlling stochastic features in image diversity and the nature of the latent space. In response, researchers have commenced exploring the inner workings of generators.

The rise of StyleGAN is in direct reaction to this challenge. It modifies the generator network structure to improve the image generation process. Using a fixed, learned input, the generator adjusts the "style" of each convolutional layer to the underlying code for more direct image feature management. Moreover, StyleGAN can somewhat modify attributes and execute style blending or interpolation operations with noise injection. This approach permits researchers to gain a more in-depth comprehension of the image generation process and control model generation style.

Dong and his colleagues [20] employed a data enhancement method based on StyleGAN in their research to generate diverse types of pavement damage images without replacing the original dataset images. This technique enhanced the segmentation model's efficacy, as confirmed by Figure 4, which shows the model's architecture.
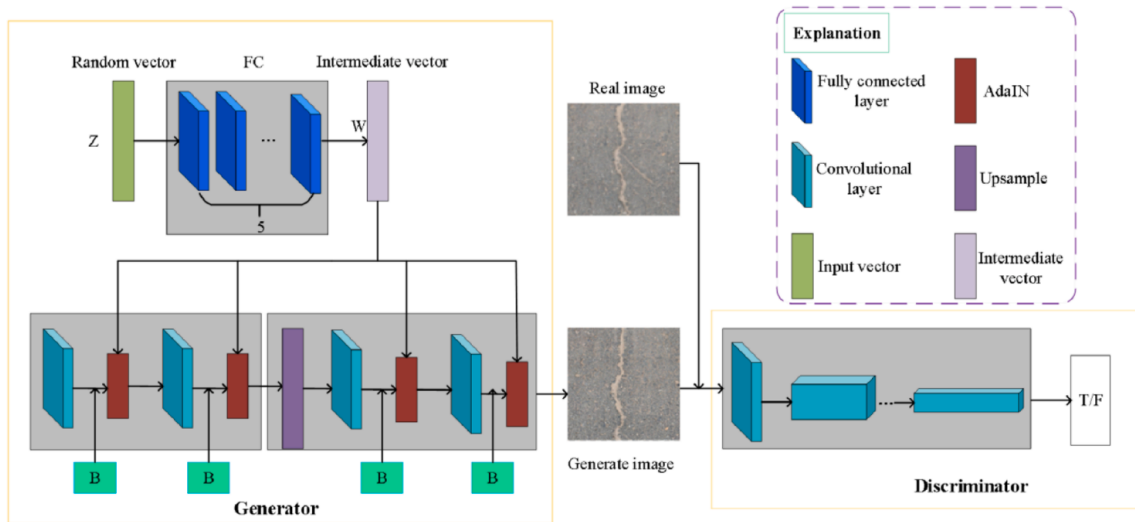
Figure 4. Pavement damage data enhancement model based on StyleGAN [20].

3. Conditional generation networks for road image data enhancement

3.1. Dataset construct and preprocess

We selected several current open-source datasets of pavement cracks for constructing the asphalt pavement crack image dataset to improve the generalization performance. The dataset used in this paper includes transverse crack and longitudinal crack images from CrackForest, Crack500, Cracktree20, GAPs384 and AEL (Aigle-RN & ESAR & LCMS).

The CrackForest dataset is a database of road crack images that respond to the condition of urban pavements under general inspection conditions [21], and there are 156 images in total, which are filtered according to the acquisition of crack characteristics. Crack500 [22] dataset was derived from 500 photos taken by cell phones at Temple University. To fit the model input image pixel requirements and to facilitate model training, the original images are divided into non-overlapping image regions.CrackTree200 contains 206 pavement images with multiple types of cracks, along with lighting, noise and low contrast [23]. GAPs384 originated from the German asphalt pavement distress dataset, which is a large sample dataset of a high standard and contains a total of 1969 grayscale images that also contain different categories of cracks and potholes [24]. A total of 509 images of cracks were selected among them. AEL is the conditioned pavement image database obtained from different travel speeds [25].

The images used for the experiments in this paper are transverse cracks (1037 images) and longitudinal cracks (1107 images) of asphalt pavements. After converting them into inputs suitable for the model, the crack images that qualified by manual selection were counted as shown in Table 1.

Furthermore, because the crack images in the selected dataset have different acquisition methods and road conditions, the images include different characteristics of background conditions and environmental conditions, and if the original images are not grayed out, many color information with low correlation to the crack morphology is learned in the subsequent network model, resulting in a serious distortion of the generated crack images. Therefore, this paper's images are grayed out to solve this issue.

Table 1. Detailed data of the selected dataset.

Dataset	CrackForest	Crack500	CrackTree200	GAPs384	AEL
Image resolution	480 × 320	2000 × 1500	800 × 600	1920 × 1080	800 × 800
Amount	96	1071	595	325	57
Total			2144		

3.2. Conditional generative adversarial network and analysis

In the unconditional Generative Adversarial Network [26] framework, the process of generating new data samples is essentially random, relying on inputs in the form of randomly generated noise vectors, as shown in Figure 2(a). This inherent randomness is valuable for generating diversity and presents significant challenges in terms of direct ability and precise control. Unlike conditional models that allow the manipulation of specific attributes, unconditional GANs do not provide explicit levers to influence the characteristics of the output, such as determining the specific type or style of the generated samples. Thus, while the model excels at generating a wide range of outputs, the lack of deterministic control mechanisms means that guiding the generative process to produce results with desired properties remains an elusive endeavor.

To address the limitations of unconditional GAN, a new version called conditional GAN has been introduced. This approach incorporates label information into the generation process, providing guidance for the model. With conditional GAN, both the generator and discriminator models are conditioned on these labels. This approach differs significantly from the traditional GAN as it relies on guidance during the generation process. Its advantage lies in its ability to produce data with specific properties. For instance, it can create images that fall within a certain category. This is a marked contrast to the arbitrary and unregulated results of unconditional GANs. Therefore, conditional GANs offer a potent means of regulating the generation process's intricacies, thereby augmenting GANs' practicality and versatility across different domains.

3.3. Case studies for conditional generative adversarial networks

3.3.1.　Case 1: Road disease data enhancement

The application of cGANs extends beyond merely enlarging pavement disease datasets; it also includes generating class-specific data to address the imbalance in data clustering classes, which is a result of the rarity of certain diseases. Integrating with labels, cGANs can effectively produce data for specific types of pavement diseases, which is particularly important for addressing the imbalance of pavement disease types. In traditional datasets, some types of diseases may have less data due to their rarity, which hinders effective learning in model training and recognition processes. Using cGANs, we can generate more data for these rare disease types, thereby enhancing the model's recognition ability and accuracy, and ensuring effective identification and handling of various types of pavement diseases. In this way, cGANs provide an effective solution for balancing and enriching pavement disease datasets. Specific cases are listed below:

Tang et al. [27] proposed a fault diagnosis method based on the Wasserstein Generative Adversarial Network (WGAN-GP) and Convolutional Neural Network (CNN) for the problem of severe imbalance and distributional differences in fault data. Ten datasets of unbalanced states were

used in the study and fed into five other deep-learning models for comparison experiments. Experimental results showed that this method achieved a fault diagnosis accuracy of 99.9% under the first data distribution condition, while the accuracies of the other six methods were 99.1, 98.9, 98.6, 97.8, 94.1 and 93.9%, respectively. When the number of training samples for each fault category was reduced to half of the normal sample count, this method's fault diagnosis accuracy far exceeded the other six methods, reaching 99.2%. We also found that as the data imbalance ratio increased, the diagnostic performance of each method significantly decreased. However, when the imbalance ratio reached 10: 1, this method exhibited good diagnostic performance. Therefore, although the fault identification accuracy of this method decreased with the increase in data imbalance ratio, it maintained high diagnostic accuracy and stability. This research provides an intelligent method for diagnosing rolling bearing faults and has proven its effectiveness and superiority in experiments.

3.3.2. Case 2: Pavement disease iinspection

The application of Conditional Generative Adversarial Networks (cGANs) is not limited to data generation. In fact, they can also be used for detecting road surface diseases. These networks, by learning various characteristics and patterns of pavement diseases, can effectively recognize and categorize different types of diseases. In the application of road disease detection, cGANs are capable of analyzing road images and identifying the presence and category of diseases. This is because cGANs can understand and mimic the characteristics of pavement diseases, thereby providing accurate identification during the actual detection process. This method is particularly effective for subtle or complex diseases that are difficult to identify through traditional methods. Thus, cGANs offer an efficient and accurate technological means for the detection and classification of road surface diseases, significantly improving the efficiency and effectiveness of pavement disease management and maintenance.

Kyslytsyna et al. [28] conducted a study on a road surface crack detection method based on Conditional Generative Adversarial Networks, proposing an automatic method to detect road surface cracks using deep learning technology. They trained a conditional generative adversarial network to generate cracks in real pavement images and compared them with actual images for the detection and localization of cracks. They mention existing crack detection methods but highlights issues with these methods in processing real-world images, such as high false-positive rates and sensitivity to lighting and shadows. To address these issues, they proposed a method based on Conditional Generative Adversarial Networks that can detect road surface cracks more accurately. The research team validated their method experimentally and compared it with other methods. The results showed that their method has high accuracy and robustness in crack detection. Additionally, they discussed the limitations and potential improvements of their approach. Overall, we present a road surface crack detection method based on Conditional Generative Adversarial Networks and demonstrates its effectiveness through experiments. This research is significant for road maintenance and safety, potentially enhancing the accuracy and efficiency of pavement crack detection.

3.3.3. Conditional Generative Adversarial Network (CGAN)

One limitation of the traditional GAN is that if the input image is ambiguous, the model produces a random image and lacks the ability to determine the image class. The CGAN [29] enhances the

original GAN by incorporating supplementary conditional data into the original network structure, resulting in a more objective function when compared to the traditional GAN, for both the discriminator and the generator input. From solving the binary game problem of minimal maximization between the generator and the discriminator, we progress to solving the minimal-maximal binary game problem with conditional probabilities. We use the CGAN model, as presented in Figure 5.

1) The Generator structure consists of four fully connected layers that require the input of conditional information and noise data, with a dimension of 100, to obtain the generated image.

2) The structure of the discriminator involves inputting condition information and image data and deriving the probability of image authenticity via four fully connected layers.
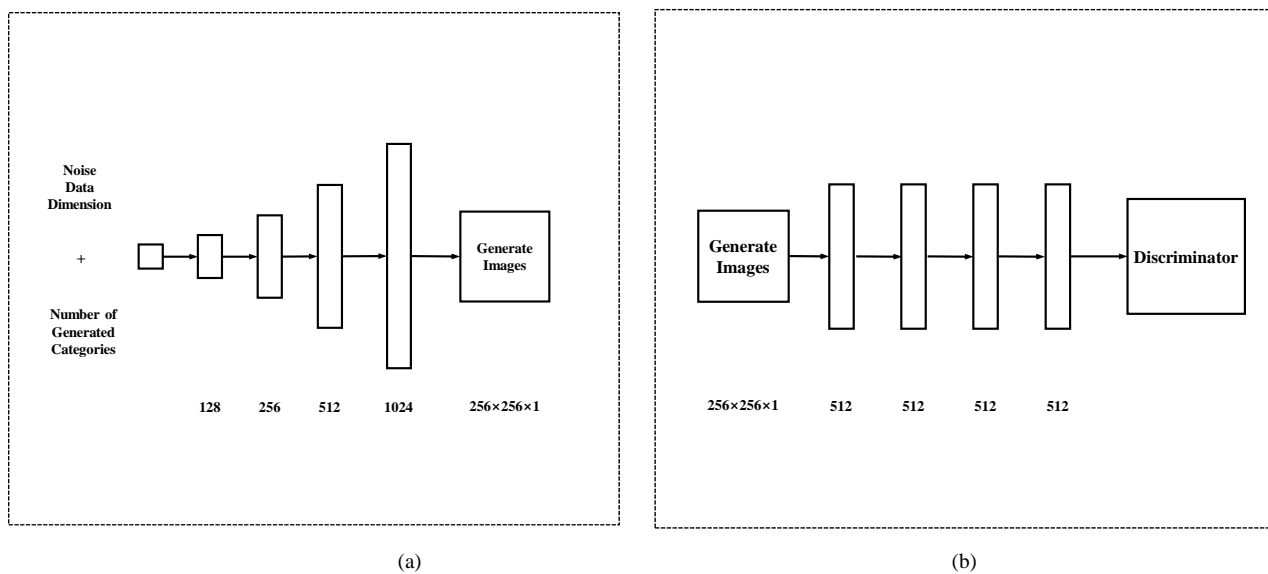
Figure 5. Conditional Generative Adversarial Network (CGAN) network structure. (a) Generator structure of CGAN. (b) Discriminator structure of CGAN.

3.3.4.   Auxiliary Classifier Generative Adversarial Network (ACGAN)

ACGAN [30] can be considered as a combination of DCGAN and CGAN. The network structure of ACGAN is shown in Figure 6. Feature extraction capability of convolution is used to improve the learning of the model.

1) Generator structure: The network has 5 layers, 4 convolutional layers and 1 fully connected layer. The input is set to random noise (variable) of 1×100 dimensions and satisfying uniform distribution. The image with a feature map of $4 \times 4 \times 1024$ is obtained by the fully connected layer; the generated crack image is obtained by 4 times transposed convolution.

2) Discriminator structure: The input of the discriminator is the generated image and the real image, and the output is the probability of judging the authenticity of the image. The feature map is reduced by 4 times of convolution operation, and then the input image is true/false by a fully connected layer.
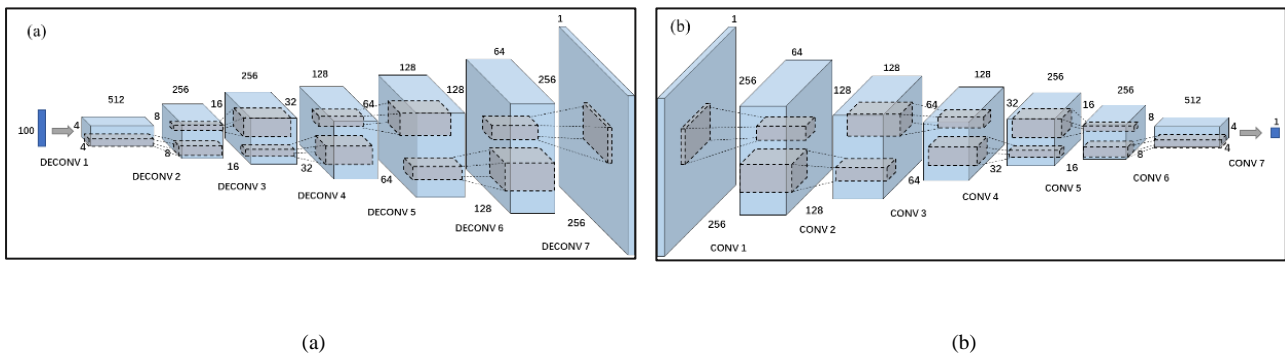
(a) (b)

Figure 6. Adversarial Conditional Generative Adversarial Network (ACGAN) network structure (a) Generator structure of ACGAN (b) Discriminator structure of ACGAN.

ACGAN makes the following improvements to the original CGAN. Eliminate all pooling layers. Except for the output layer of the generative network and the input layer of the discriminative network, both use BN (Batch Normalization) to handle the initialization problem. Eliminate the fully connected layer and transpose the 1×1 convolution to make the network a fully convolutional network. ReLU is used as the activation function in the generative network, tanh is used in the last layer and LeakyReLU is used as the generating function in the discriminative network.

3.3.5. Self-Attention Generative Adversarial Network (SAGAN)

Traditional convolutional GAN generates images very realistically on texture features but does not work well in generating geometric structure features. Due to the fixed size of the convolutional kernel in the convolutional neural network, only local features can be extracted, and geometric features that depend on global features may not be extracted when they appear. One method is to expand the perceptual field by increasing the number of layers of the network, which leads to an increment in model depth and training efficiency. Another approach is to increase the size of the convolutional kernel to enhance the perceptual field, but this approach increases the number of parameters and the training cost. For this reason, Zhang et al. [6] applied the self-attentive mechanism, which has been widely used in recent years, to obtain global features and thus enhance the geometric features of the generated images with less training cost.

The principle of the attention mechanism is shown in Figure 7 below, where Q (query), K (key) and V (value) are selected, and the probability distribution is obtained by matrix multiplication of Q and K by Softmax, which is then multiplied by V to obtain the self-attention value.
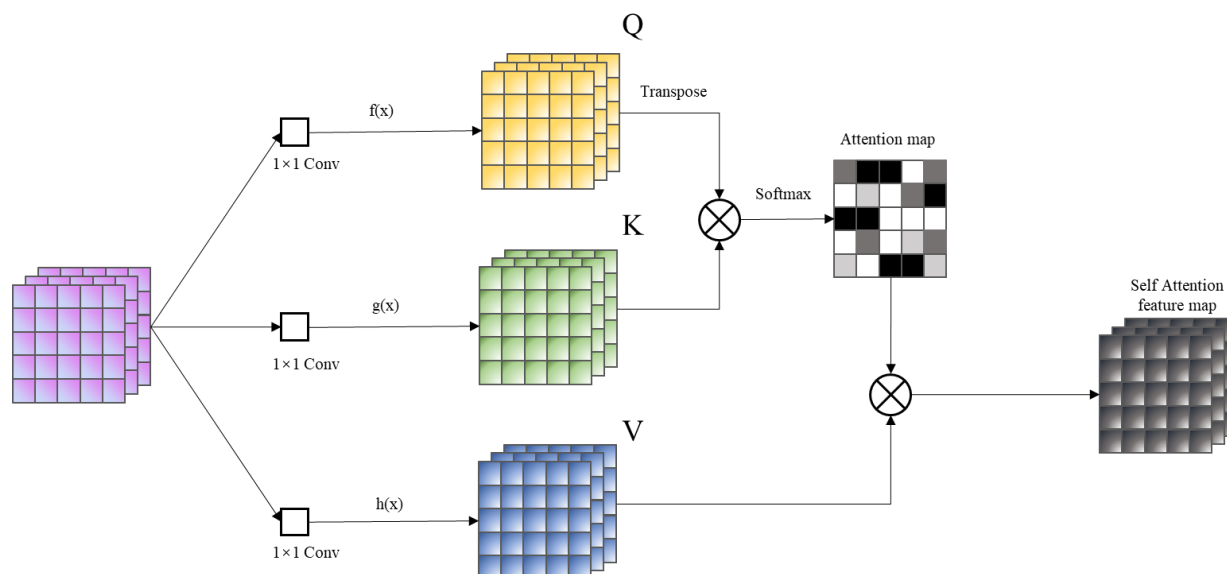
Figure 7. Schematic diagram of self-attention in Self-Attention Generative Adversarial Network (SAGAN).

The self-attention mechanism implies that Q, K and V all originate from themselves and can be updated based on task training. Self-attentions are applied to the latter two layers of both discriminator and generator.

The network architecture of the generator consists of the following components: First, the generator receives a random noise vector and labels it as input. It uses a fractionally-strided convolution for up-sampling, combined with batch normalization and a ReLU activation function. This noise vector is converted into a larger feature map in multiple up-sampling layers, each using transposed convolution, batch normalization and the ReLU activation function. In addition, the generator contains two Self-Attention layers that help the model capture long-range dependencies in the image.

The discriminator part is initialized with down-sampling using ordinary convolutional layers with a LeakyReLU activation function. The features of the input image are progressively compressed in multiple down-sampling layers, each using convolution and LeakyReLU. The discriminator also contains two self-attention layers to improve the recognition of image details. Finally, the final layer of the discriminator uses convolution to convert the feature map into a score indicating the probability that the image is real.

Spectral normalization [6] is added for the generator and discriminator, which effectively reduces the computational effort of training and makes it more stable. The SAGAN network structure is shown in Figure 8 below.
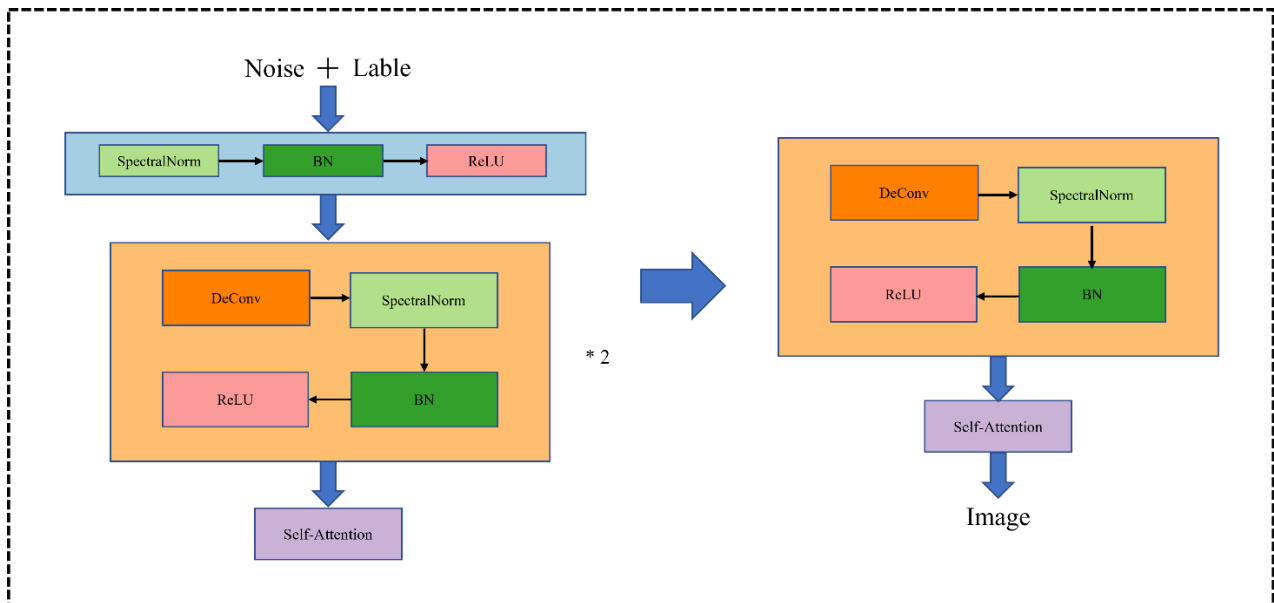
(a)
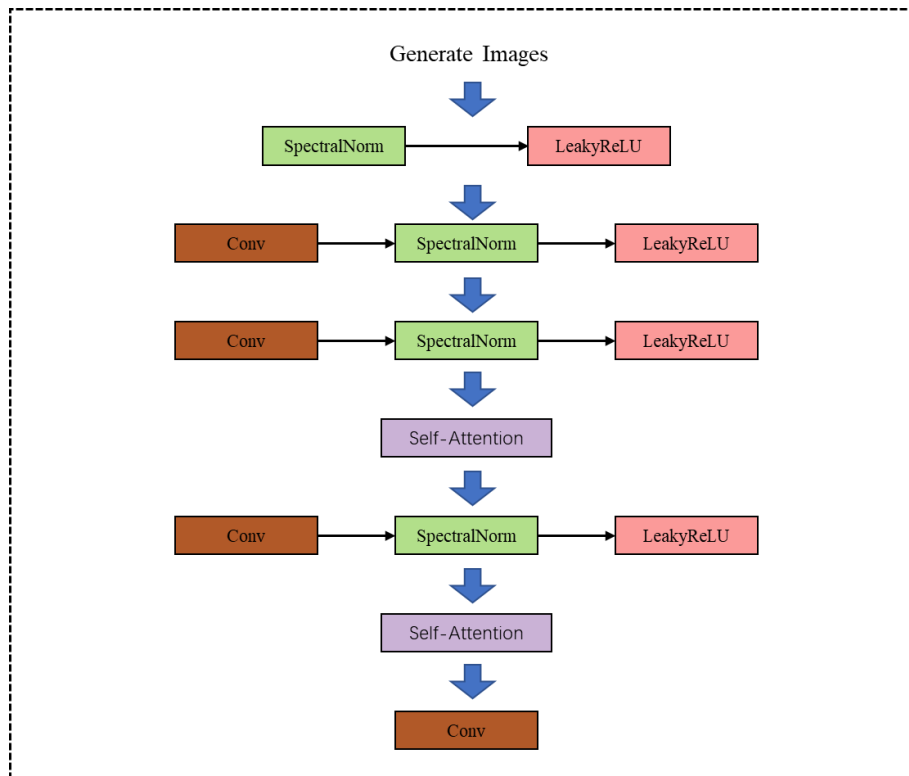
(b)

Figure 8. Self-Attention Generative Adversarial Network (SAGAN) network structure.
(a) Generator structure of SAGAN. (b) Discriminator structure of SAGAN.

In addition to the self-attention mechanism, the SAGAN in this paper applies Spectral Normalization to the weights of the generator and discriminator based on ACGAN, so that the discriminator satisfies the 1-Lipschitz condition, and avoids the gradient anomalies caused by too many parameters of the generator, making the whole training process more stable.

4. Results

4.1. Image generation results

Images of asphalt pavement cracks generated by CGAN, ACGAN and SAGAN on a different number of iterations are shown in Figure 9 CGAN, ACGAN and SAGAN all use SGD optimizers with a generator learning rate of 0.0001, a discriminator learning rate of 0.0004 and a batch size uniformly set to 16. Among the images generated by each GAN model, the generated images of transverse cracks and longitudinal cracks are randomly selected at each iteration number for comparison. CGAN, ACGAN and SAGAN are shown with 100, 50 and 20 epochs as 1 stage, respectively.

The evaluation of the image generation effect was studied in this paper. There is currently no fixed and objective metric for the evaluation of images generated by conditional GAN, and the objective evaluation of the model generation effect is performed using common metrics in previous studies. Namely, the Inception Score (IS) and the FID (Fréchet Inception Distance) metrics for the original crack dataset used in this paper as well as the generated ones. The IS index is usually used to evaluate the GAN-generated images.

IS is usually used to evaluate the quality and diversity of the GAN-generated images using the score obtained on Inception Net-V3, and a high score usually indicates that the generated images meet the requirements. The Inception Score (IS) is a metric used to evaluate the quality of generated images. Its formula is defined as follows:

$$IS(x) = exp^{DKL(p(y|x)||p(y))} \tag{3}$$

x represents the generated image, $p(y|x)$ denotes the probability distribution of the predicted class y given the image x and the model, and p(y) is the marginal distribution of the predicted class y. The IS determines the quality of the generated image by computing the DKL divergence between the conditional class distribution $p(y|x)$ and the marginal distribution $p(y)$. If the generated image entails discernible objects, the model's conditional class distribution output is likely to exhibit low entropy and proximity to the edge distribution. This, in turn, leads to a high IS score.

The FID distance score is a measure to calculate the distance between the feature vectors between the features of the real image and the generated image. It is calculated using the Inception v3 image classification model, and a low score means that the two sets of images are more similar, i.e., the GAN generates images that are similar to the real images. FID (Fréchet Inception Distance) [31] is another measure of the quality of the generated image, which is calculated as.

$$d^2(F, G) = |\mu_X - \mu_Y|^2 + tr[\sum X + \sum_Y Y - 2(\sum X \sum Y)^{\frac{1}{2}}] \tag{4}$$

where X and Y represent the distribution of the generated image and the real image respectively, μ and Σ are the mean and covariance matrices of the features extracted by the Inception-v3 [32] model, respectively. FID evaluates the quality of the generated image by calculating the statistical difference between the mean and covariance matrices of the features of the two, and the lower the score of FID, the lower the visual differentiation between the generated image and the real image. Where x and y represent the distribution of the generated image and the real image, respectively. μ and Σ denote the mean and covariance matrices of the features extracted by the Inception-v3 model. The quality of the generated image is evaluated by FID, which calculates the statistical difference between the mean and covariance matrices of the features of both images. The lower the score of FID, the lower the visual

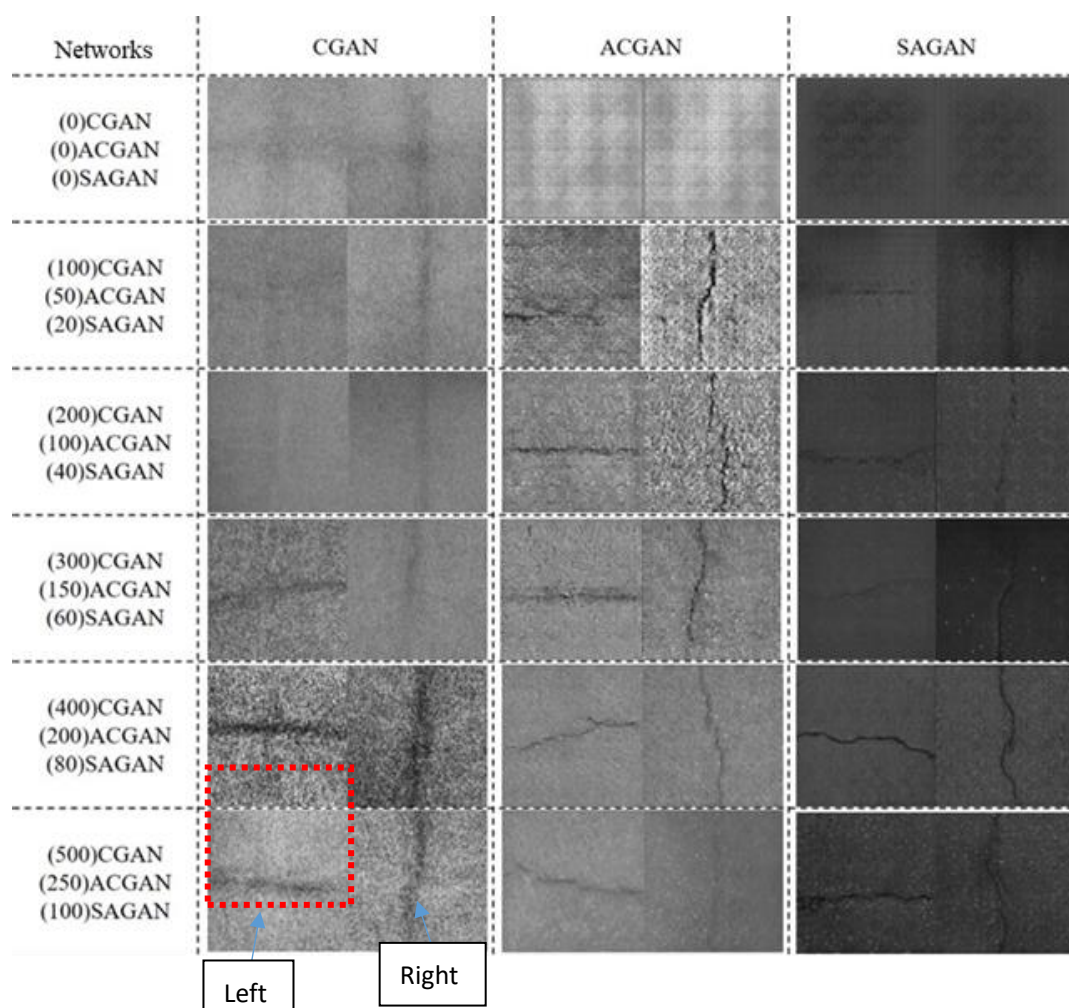differentiation between the generated and real image.

Figure 9. Different GAN iterations generate images (several iterations are indicated in parentheses): The left side of the image represents the horizontal crack, while the right side represents the vertical crack.

From Figure 9, it can be seen that when iterating to stage 1, CGAN just has traces of cracks, ACGAN appears as the outline of cracks and SAGAN appears as the location of cracks, all of which are in a serious and distorted degree of meshing at this time. When the iteration is at stage 2, CGAN does not show the crack outline, ACGAN, and SAGAN show the crack outline and the meshing problem is improving. When iterating at stage 3, CGAN cracks are still not ideal, ACGAN and SAGAN crack outlines are clearer, and the gridding problem is greatly improved. when iterating at stage 4, CGAN shows crack outlines, ACGAN and SAGAN generate images with slightly blurred backgrounds, but the crack theory opening has taken shape. at stage 5, CGAN starts to involve the difference between background and crack. At stage 5, CGAN starts to involve the difference between the background and the crack.

Table 2. Comparison of different gan evaluation metrics.

Type of cracks	CGAN		ACGAN		SAGAN	
	Longitudinal Crack	Transverse Crack	Longitudinal Crack	Transverse Crack	Longitudinal Crack	Transverse Crack
IS	1.44 ± 0.03	1.33 ± 0.06	1.61 ± 0.04	1.46 ± 0.05	1.73 ± 0.05	1.59 ± 0.07
FID	133.76	124.06	81.15	73.73	71.29	68.01

5. Discussion

The evaluation of the image generation effect was studied in this paper. The Inception Score (IS) values for longitudinal and transverse cracks in the original images are 2.45 ± 0.11 and 1.98 ± 0.09, respectively. Data from Table 2 shows that SAGAN significantly outperforms both CGAN and ACGAN in generating images of longitudinal and transverse cracks. The high IS values (1.73 ± 0.05 and 1.59 ± 0.07) of SAGAN, along with its low Frechet Inception Distance (FID) values (71.29 and 68.01), suggest its superior capability in rendering more realistic and diverse crack images.

This superiority is likely attributed to SAGAN's implementation of a self-attention mechanism, which enhances its ability to process global information and manage long-range dependencies in images, a critical aspect of pavement crack analysis. By analyzing the experimental results, we found that the generative performance advantages of SAGAN have great potential to be a beneficial tool, assisting road engineers in high-quality construction and more balanced pavement disease detection datasets.

The analysis also reveals a consistent pattern across all tested conditional GANs: IS values for longitudinal cracks are higher than those for transverse cracks, while FID values show an inverse relationship. This indicates a general trend in the models' capabilities, where longitudinal cracks are rendered with high quality and diversity, more closely resembling real images. Such a trend might stem from the intrinsic differences in the complexity or visual features of longitudinal versus transverse cracks.

While these results confirm the effectiveness of SAGAN in pavement crack image processing, they also open avenues for further research. It is crucial to delve deeper into the reasons why SAGAN outperforms its counterparts. Factors such as the quantity and quality of training data, the architectural nuances of the model and the specifics of the training process can influence the performance. Moreover, aligning these findings with existing theories in image processing and GANs can provide valuable insights into the model's behavior.

For practical deployment, understanding these dynamics is crucial. It can guide the development of more refined models for pavement crack detection, contributing to more efficient and accurate maintenance strategies in urban planning and infrastructure management. Future research directions might focus on experimenting with varied training datasets, tweaking model architectures and exploring hybrid models that combine the strengths of different GANs to enhance image generation quality for specific types of pavement cracks.
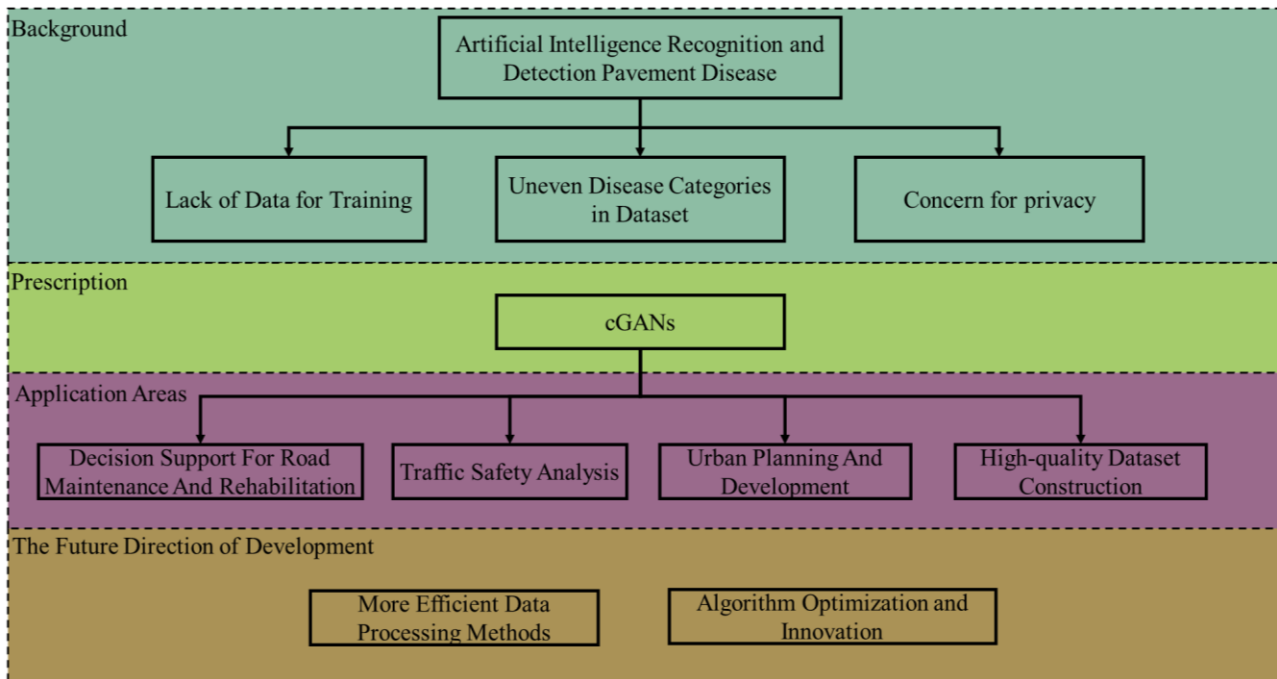
Figure 10. Future research directions and practical application roadmap in the field of pavement crack image generation.

6. Conclusions

For the limited number of current pavement-specific field crack images, we carried out the optimal design of the GAN model from crack dataset augmentation using public datasets.

1) The GAN was designed based on the structure of fully connected, convolutional and attention mechanisms, and it was demonstrated subjectively and objectively that the use of attention mechanisms can improve the quality and diversity of the images generated by the GAN, which are more similar in content to the original crack image dataset.

2) By integrating the structure of CGAN within SAGAN and its self-attention mechanisms, this method has partially solved the problem of GANs being unable to generate category-specific images, while also enhancing the quality of image rendering. The approach presented in this paper is not only applicable for expanding the pavement disease image dataset but also addresses the issue of imbalanced class samples within the dataset. This enhancement in dataset quality ensures the accuracy of subsequent identification and detection tasks.

The results demonstrate that the conditional GAN with the self-attention mechanism produces high-quality pavement crack images, such as increased detail, clarity and realism. This improvement in image quality is a significant advantage and can potentially benefit applications related to road maintenance and safety. However, the self-attention mechanism demands considerable computational resources. To optimize and accelerate the model's performance, we plan to simplify the network structure, refine parameters and compress the model.

Use of AI tools declaration

The authors declare that they have not used artificial intelligence tools in the creation of this article.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. X. Guan, H. Zhang, X. Du, X. Zhang, M. Sun, Y. Bi, Optimization for asphalt pavement maintenance plans at network level: Integrating maintenance funds, pavement performance, road users, and environment, *Appl. Sci.*, **13** (2023), 8842. https://doi.org/10.3390/app13158842

2. Y. Du, N. Pan, Z. Xu, F. Deng, Y. Shen, H. Kang, Pavement distress detection and classification based on YOLO network, *Int. J. Pavement Eng.*, **22** (2021), 1659–1672. https://doi.org/10.1080/10298436.2020.1714047

3. J. Zhu, J. Zhong, T. Ma, X. Huang, W. Zhang, Y. Zhou, Pavement distress detection using convolutional neural networks with images captured via UAV, *Autom. Constr.*, **133** (2022), 103991. https://doi.org/10.1016/j.autcon.2021.103991

4. E. Ibragimov, H. J. Lee, J. J. Lee, N. Kim, Automated pavement distress detection using region based convolutional neural networks, *Int. J. Pavement Eng.*, **23** (2022), 1981–1992. https://doi.org/10.1080/10298436.2020.1833204

5. J. Guan, X. Yang, L. Ding, X. Cheng, V. C. Lee, C. Jin, Automated pixel-level pavement distress detection based on stereo vision and deep learning, *Autom. Constr.*, **129** (2021), 103788. https://doi.org/10.1016/j.autcon.2021.103788

6. H. Zhang, I. Goodfellow, D. Metaxas, A. Odena, Self-attention generative adversarial networks, in *Proceedings of the 36th International Conference on Machine Learning*, **97** (2019),7354–7363. https://doi.org/10.48550/arXiv.1805.08318

7. Z. Tong, T. Ma, W. Zhang, J. Huyan, Evidential transformer for pavement distress segmentation, *Comput. Aided Civ. Infrastruct. Eng.*, **38** (2023), 2317–2338. https://doi.org/10.1111/mice.13018

8. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial networks, *Adv. Neural Inf. Process. Syst.*, **27** (2020). https://doi.org/10.48550/arXiv.1406.2661

9. M. Arjovsky, S. Chintala, L. Bottou,Wasserstein generative adversarial networks, *Int. Conf. Mach. Learn.*, (2017), 214–223. https://doi.org/10.48550/arXiv.1701.07875

10. T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of gans for improved quality, stability, and variation, preprint, arXiv:1710.10196. https://doi.org/10.48550/arXiv.1710.10196

11. X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, S. P. Smolley, Least squares generative adversarial networks, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, (2017), 2794–2802. https://doi.org/10.48550/arXiv.1611.04076

12. T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019), 4401–4410. https://doi.org/10.48550/arXiv.1812.04948

13. L. Pei, Z. Sun, L. Xiao, W. Li, J. Sun, H. Zhang, Virtual generation of pavement crack images based on improved deep convolutional generative adversarial network, *Eng. Appl. Artif. Intell.*, **104** (2021), 104376. https://doi.org/10.1016/j.engappai.2021.104376

14. B. Xu, C. Liu, Pavement crack detection algorithm based on generative adversarial network and convolutional neural network under small samples, *Measurement*, **196** (2022), 111219. https://doi.org/10.1016/j.measurement.2022.111219

15. D. Mazzini, P. Napoletano, F. Piccoli, R. Schettini, A novel approach to data augmentation for pavement distress segmentation, *Comput. Ind.*, **121** (2020). https://doi.org/10.1016/j.compind.2020.103225

16. L. L. Pei, Z. Y. Sun, L. Y. Xiao, W. Li, J. Sun, H. Zhang, Virtual generation of pavement crack images based on improved deep convolutional generative adversarial network, *Eng. Appl. Artif. Intell.*, **104** (2021). https://doi.org/10.1016/j.engappai.2021.104376

17. I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. C. Courville, Improved training of wasserstein gans, *Adv. Neural Inf. Process. Syst.*, **30** (2017). https://doi.org/10.48550/arXiv.1704.00028

18. K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, Ieee: 'Deep residual learning for image recognition', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 770–778

19. Z. Xu, X. Yu, Z. Liu, S. Zhang, Q. Sun, N. Chen, et al., Safety monitoring of transportation infrastructure foundation: Intelligent recognition of subgrade distresses based on B-Scan GPR images, *IEEE Trans. Intell. Transp. Syst.*, (2022), 15468–15477. https://doi.org/10.1109/TITS.2022.3224769

20. J. X. Dong, N. N. Wang, H. Y. Fang, Q. F. Hu, C. Zhang, B. S. Ma, et al., Innovative method for pavement multiple damages segmentation and measurement by the Road-Seg-CapsNet of feature fusion, *Constr. Build. Mater.*, **324** (2022). https://doi.org/10.1016/j.conbuildmat.2022.126719

21. Y. Shi, L. M. Cui, Z. Q. Qi, F. Meng, Z. S. Chen, Automatic road crack detection using random structured forests, *IEEE Trans. Intell. Transp. Syst.*, **17** (2016), 3434–3445. https://doi.org/10.1109/tits.2016.2552248

22. F. Yang, L. Zhang, S. J. Yu, D. Prokhorov, X. Mei, H. B. Ling, Feature pyramid and hierarchical boosting network for pavement crack detection, *IEEE Trans. Intell. Transp. Syst.*, **21** (2020), 1525–1535. https://doi.org/10.1109/tits.2019.2910595

23. Q. Zou, Y. Cao, Q. Q. Li, Q. Z. Mao, S. Wang, Crack Tree: Automatic crack detection from pavement images, *Pattern Recognit. Lett.*, **33** (2012), 227–238. https://doi.org/10.1016/j.patrec.2011.11.004

24. M. Eisenbach, R. Stricker, D. Seichter, K. Amende, K. Debes, M. Sesselmann, et al., How to get pavement distress detection ready for deep learning? A systematic approach, in *2017 International Joint Conference on Neural Networks (IJCNN)*, (2017), 2039–2047. https://doi.org/10.1109/IJCNN.2017.7966101

25. R. Amhaz, S. Chambon, J. Idier, V. Baltazart, Automatic crack detection on two-dimensional pavement images: An algorithm based on minimal path selection, *IEEE Trans. Intell. Transp. Syst.*, **17** (2016), 2718–2729.https://doi.org/10.1109/TITS.2015.2477675

26. A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, preprint, arXiv:1511.06434. https://doi.org/10.48550/arXiv.1511.06434

27. H. Tang, S. Gao, L. Wang, X. Li, B. Li, S. Pang, A novel intelligent fault diagnosis method for rolling bearings based on Wasserstein generative adversarial network and Convolutional Neural Network under Unbalanced Dataset, *Sensors*, **21** (2021), 6754. https://doi.org/10.3390/s21206754

28. A. Kyslytsyna, K. Xia, A. Kislitsyn, I. Abd El Kader, Y. Wu, Road surface crack detection method based on conditional generative adversarial networks, *Sensors*, **21** (2021), 7405. https://doi.org/10.3390/s21217405

29. M. Mirza, S. Osindero, Conditional generative adversarial nets, preprint, arXiv:1411.1784. https://doi.org/10.48550/arXiv.1411.1784

30. A. Odena, C. Olah, J. Shlens, Conditional image synthesis with auxiliary classifier gans, *Int. Conf. Mach. Learn.*, (2017), 2642–2651. https://doi.org/10.48550/arXiv.1610.09585

31. D. C. Dowson, B. V. Landau,The Fréchet distance between multivariate normal distributions, *J. Multivar. Anal.*, **12** (1982), 450–455. https://doi.org/10.1016/0047-259X(82)90077-X

32. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 2818–2826. https://doi.org/10.48550/arXiv.1512.00567