Article

Data Augmentation by an Additional Self-Supervised CycleGAN-Based for Shadowed Pavement Detection

Jiajun Song [1], Peigen Li [1], Qiang Fang [1], Haiting Xia [1,2,*] and Rongxin Guo [1]

1   Yunnan Key Laboratory of Disaster Reduction in Civil Engineering, Faculty of Civil Engineering and Mechanics, Kunming University of Science and Technology, Kunming 650500, China

2   Faculty of Civil Aviation and Aeronautics, Kunming University of Science and Technology, Kunming 650500, China

*   Correspondence: haiting.xia@kust.edu.cn

Abstract: With the rapid development of deep learning, pavement crack detection has started to shift from traditional manual visual inspection to automated detection; however, automatic detection is still a challenge due to many complex interference conditions on pavements. To solve the problem of shadow interference in pavement crack detection, this paper proposes an improved shadow generation network, named Texture Self-Supervised CycleGAN (CycleGAN-TSS), which can improve the effect of generation and can be used to augment the band of shadowed images of pavement cracks. We selected various images from three public datasets, namely Crack500, cracktree200, and CFD, to create shadowed pavement-crack images and fed them into CycleGAN-TSS for training to inspect the generation effect of the network. To verify the effect of the proposed method on crack segmentation with shadow interference, the segmentation results of the augmented dataset were compared with those of the original dataset, using the U-Net. The results show that the segmentation network achieved a higher crack recognition accuracy after the augmented dataset was used for training. Our method, which involves generating shadowed images to augment the dataset and putting them into the training network, can effectively improve the anti-shadow interference ability of the crack segmentation network. The research in this paper also provides a feasible method for improving detection accuracy under other interference conditions in future pavement recognition work.

Keywords: CycleGANs; shadow interference; data augmentation; crack detection

1. Introduction

With the rapid development of highway transportation, the demand for pavement maintenance is growing. Accurate and fast pavement inspection technology can provide decision support for pavement maintenance management. However, the most widely used pavement inspection method is still traditional manual inspection, which has obvious disadvantages, such as strong subjectivity and low efficiency. With the development of deep learning in recent years, certain traditional computer vision tasks have started to be performed by deep learning, such as image classification [1–3], target detection [4–6], and semantic segmentation [7–9]. In other fields, such as material design [10] and structure design [11,12] tasks, deep learning has also shown great potential. In this context, automated pavement detection has also started to receive more attention [13]. Compared with traditional detection algorithms, researchers [14] found that convolutional neural networks exhibit better performance in pavement crack recognition. Ji A et al. [15] proposed an integrated method for crack detection based on the convolutional neural network DeepLabv3+, which proved to be effective and reliable in automatic asphalt pavement inspection. Peigen Li et al. [16] proposed a crack segmentation U-Net network with edge detection output to improve the accuracy of pavement crack detection.

In the pavement inspection task, cracks need to be accurately detected and segmented. For pavement cracks in ideal conditions, traditional image-processing techniques such as

edge detection and threshold segmentation can effectively segment cracks in images [17,18]. However, in real-world pavements, crack segmentation is often affected by different pavement conditions, such as shadows and stains [19,20], with shadows representing the main environmental interference factor in the pavement crack segmentation task. In the process of automatic pavement detection, the network can misjudge and miss cracks due to the similarity between the color features of shadows and structural features and the corresponding features of cracks, thus making the detection of cracks under shadow interference problematic.

To address the interference of shadows in crack detection, there are two direct practical approaches to shadow removal: removing the shadows that create interference in crack images by using traditional image algorithms [21] and carrying out shadow removal by using deep-learning methods [22]. For deep learning, the final training effect of the network is largely related to the training dataset [23]. Therefore, the shadow-removal task for images of pavement cracks is essentially identical to the crack detection segmentation task under shadow interference, as both need to first distinguish shadows from cracks in the image. The difference lies in the fact that, in the former method, the shadowed areas need to be selected and restored to a shadow-free image. This additional task often requires a larger corresponding dataset from the trainer. Thus, with the clear need for crack segmentation, this paper adopts the latter, more straightforward approach, which was used to train our network under shadow interference so that it was able to complete crack segmentation under interference conditions.

As is well-known, deep learning is a data-driven method, and its training effect requires a sufficient amount of relevant data as support. Moreover, because of the domain sensitivity of deep learning, the datasets obtained in different situations or different periods are not universal. In practical situations, it is a challenge to collect a sufficient number of images with both cracks and shadows for network training due to the influence of weather and lighting. To improve the detection effect of pavement cracks, data augmentation is the best solution when there are not any sufficient image data.

To augment and generate more data from existing datasets, Nanni L et al. [24] proposed a new data augmentation method to construct a new set by combining 14 traditional data augmentation methods to generate new images in the dataset. However, traditional image data augmentation methods, such as rotation, flipping, stretching, etc., often cannot process specific semantic information in the image, and the augmented image data generated in this way cannot provide the network with new semantic information for learning; it can only improve the adaptability of the network to the location of image semantic information and shape distortion. Various researchers [25] proposed that more effective data can be generated from existing data through generative adversarial networks.

Goodfellow et al. [26] proposed the Generative Adversarial Network (GAN), which was trained based on a mutual game between the generator and the discriminator, and limited the iterative convergence of the network through the adversarial generation loss to generate the target image from the noise space. Phillip Isola et al. [27] proposed the Pix2Pix network, which used skip-connected U-Net as the generator and PatchGAN, which was used to capture the regional feature information, as the discriminator. In this way, the network could generate the desired image effect without destroying the basic information of the original image. To use unpaired datasets for training, Zhu J Y et al. [28] designed the Cycle-Constraint Adversarial Network (CycleGAN); similar to Pix2Pix, the CycleGAN also uses the special U-Net with skip-connected as the generator and PatchGAN as the discriminator. Since it was proposed, as a result of the fact that it can complete the transformation task of two image domains without the need for paired datasets, the network can complete many generation tasks that cannot obtain paired datasets [29,30] and complete many semantic transformation tasks with its excellent image "translation" ability [31,32]. Therefore, based on the CycleGAN, researchers [33,34] proposed CycleGANs with additional information supervision to be applied to the target generation task, which uses additional labels to strengthen the network's generation of image-specific information. However, to improve the effect of texture generation by adding supervision, manual labels

are often needed. Making these labels makes training data more time-consuming and laborious. Moreover, when the defect texture is strongly supervised by the labels, although the ideal texture is generated, the diversity of the texture depends on the number of image labels, and this creates a huge limitation for certain tasks when generating images and augmenting the dataset.

Compared with generating the target image from the noise space, using the CycleGAN to generate the semantic information of shadows requires less training and smaller datasets. On the other hand, for paired datasets in the pavement image generation task, it is difficult to obtain the same pavement images, both with and without shadows. Therefore, to complete the pavement image generation task studied in this paper, it is the best choice to adopt and transform CycleGAN for shadow generation. In order to make CycleGAN coordinate the shadow texture generation task and avoid the problems caused by enhanced supervision, based on the original CycleGAN, this paper proposes an additional self-supervised CycleGAN (CycleGAN-TSS). By adding a texture self-supervised structure to constrain the generation of shadows, the network achieves a better generation effect under self-supervision. According to the structural characteristics of the network, it is named the texture self-supervised CycleGAN (CycleGAN-TSS) network. We used this new generative network to augment the data and verify the practical effect of augmenting the image dataset.

The following are the main contributions of this paper:

- A CycleGAN for texture self-supervised for texture generation was established. This is able to generate shadow textures in a self-supervised manner, which enables the network model to output images with shadow textures and binary images with shadow information in a more stable manner;
- In this study, we conducted experiments on the crack500, cracktree200, and CFD datasets to verify the shadow-generation effect of the CycleGAN-TSS network by visual comparison and demonstrated the stability of the network for image shadow information generation on unpaired datasets and small datasets;
- It is demonstrated that the proposed shadow data augmentation method can be used to improve the anti-interference capability of pavement crack detection under the influence of shadows.

The rest of the paper is organized as follows. Section 2 describes the research methodology in detail. Section 3 describes the experimental procedure and the results. Finally, the paper is summarized in Section 4.

2. Methodology

This section first introduces the production method of the image dataset. Then it details the improvements to the original CycleGAN, i.e., the designed texture self-supervised CycleGAN (CycleGAN-TSS) network structure, which adds a new lateral shadow texture self-supervised loss function and adjusts the balance of multiple loss functions to derive the overall loss function.

2.1. Data Sources and Data Processing

The data required for training included two types of sample data: pavement crack images with shadows and pavement crack images without shadows. These two types of sample data were mainly sourced from the public pavement crack image datasets CRACK500 [35,36], CFD [37], and CrackTree200 [19]. The CRACK500 dataset contains about 500 original images of size 3264 × 2448 and a total of 3364 images after clipping, the CFD dataset contains 155 original images, and the CrackTree200 dataset contains 206 original images. The statistics of the number of images selected from the above three datasets are shown in Table 1. The image is selected to add shadow information, and the processed image is fed into CycleGAN-TSS for training. In addition to the need for shadowed-image augmentation, the CRACK500 dataset also selects various additional images to verify the actual crack-segmentation effect of U-Net after the augmentation.

Table 1. Selection of dataset and number of synthetic images.

Dataset	Original Images	Selected Images	Composite Shadow Images
CRACK500	3364	780	80
CFD	155	126	80
cracktree200	206	206	80

The pavement images with obvious cracks were selected from the three datasets for augmentation, and these image data were cropped and resized. Finally, the size of all images was unified to 256 × 256. To obtain the shadowed-pavement-crack-image data with the same image style as the original dataset, we randomly added shadow masks to the original image to generate images. By randomly cropping some areas of the shadow image, the shadow mask of the pavement image was added to the original pavement image to form the shadow mask of the image, as shown in Figure 1. In this way, we obtained three synthetic shadowed-image datasets.
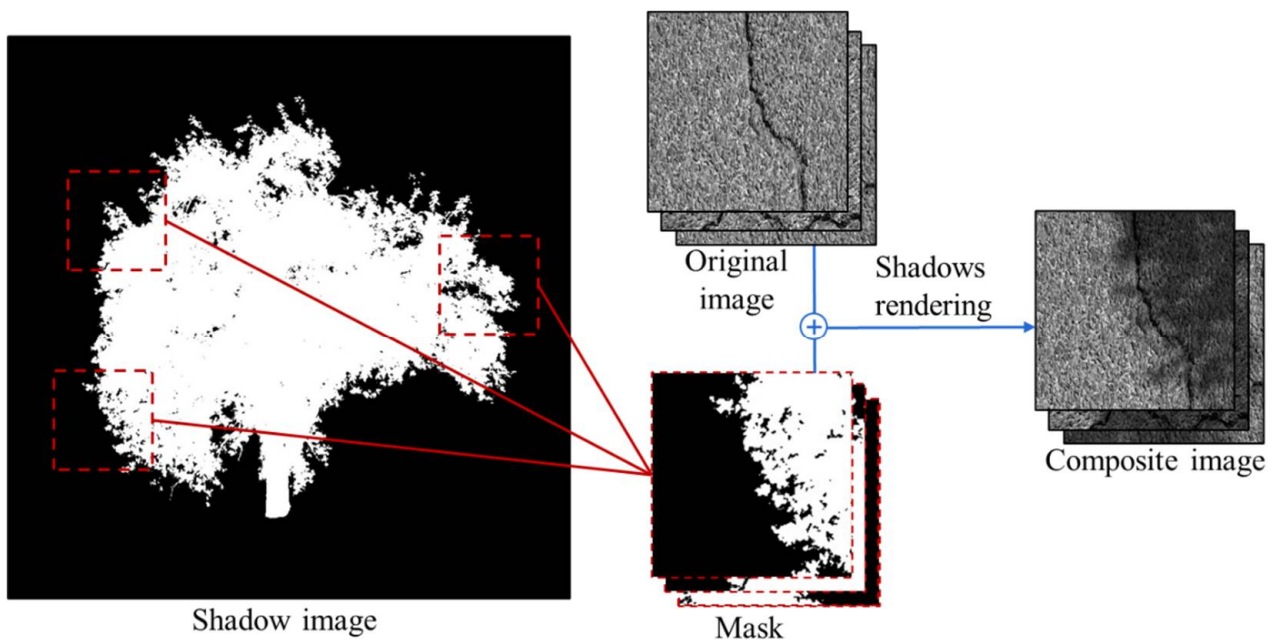
Figure 1. Illustration of making shadowed images.

2.2. Network Structure

As shown in Figure 2, the proposed CycleGAN-TSS for generating images of pavement cracks with shadows, similar to the original CycleGAN, also comes with two generators that convert between different image domains, as well as discriminators for each of the two image domains. Different from the original network, the CycleGAN-TSS comes with the generation of binary maps with shadow information when the two generators generate images. In this special-generation task with cracks and shadows, these pairwise binary maps in the network play a role in strengthening the supervision of shadow generation.
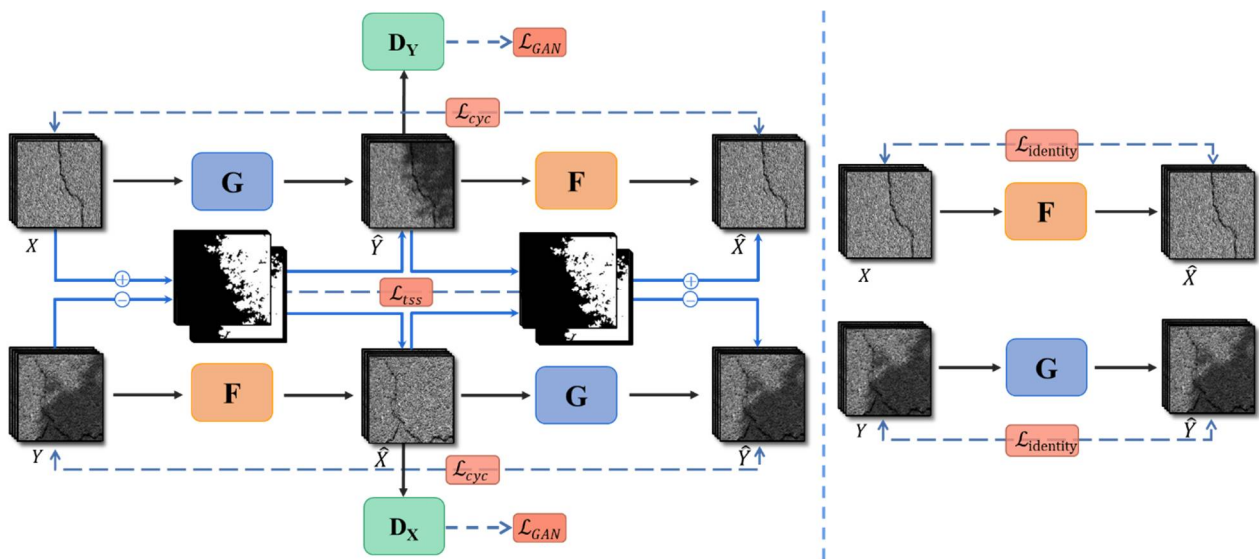
Figure 2. Schematic structure of CycleGAN-TSS. In the figure, the generator *G* is used to generate shadowed images, the generator *F* is used to generate the unshadowed images, and two discriminators D_X and D_Y distinguish between real and synthetic images, $\mathcal{L}_{GAN}$, $\mathcal{L}_{cyc}$, $\mathcal{L}_{tss}$ and $\mathcal{L}_{identity}$ are the losses in the network structure.

2.3. Loss Function

The network structure of CycleGAN-TSS proposed in this paper mainly contains two kinds of data distributions: unshadowed crack images $X \sim P_{data}(x)$ and shadowed crack images $Y \sim P_{data}(y)$. For the shadow texture image, it can be generally considered as the difference between the image semantic information in the *Y* domain and the image semantic information in the *X* domain; therefore, we can express the shadow semantic information *M* as follows:

$$Y = X + M. \tag{1}$$

This is interpreted as the *Y*-domain image semantics being equivalent to the sum of *X*-domain image semantics and shadow semantics. However, this relationship is not a simple additive relationship on image processing, and it needs a more realistic simulation effect of shadow covering the pavement. Therefore, we needed to train the generative network model by using deep learning to accomplish the task of adding shadow semantic information to *X*-domain images. Normally, CycleGAN can accomplish the transformation of semantic information to achieve the generation of image local effects, but in this particular pavement-crack-image shadow-generation task, the original CycleGAN suffers from two main problems:

- The semantic information of images with shadows and images without shadows is asymmetric. The image generation in this task requires more generation of semantic information rather than conversion of semantic information. If the original CycleGAN is used to complete this work, it can lead to color loss and structural damage to crack information in the image;
- When the original CycleGAN performs the generation task, the training images are unpaired. As a result of the lack of constraints on the recognition of shadow information, under the interference of the original crack color information, the authenticity of the generated shadow information is poor, and the image information will also be lost.

As shown in Figure 3, in the case of a small dataset, the original CycleGAN often suffers from the loss of crack color information (shown in (a)) or the loss of basic image information generation (shown in (b)) at the early stage of training. Moreover, it also suffers from overfitting caused by the small dataset (shown in (c)) and the loss of crack structure information generation (shown in (d)) during the training process.
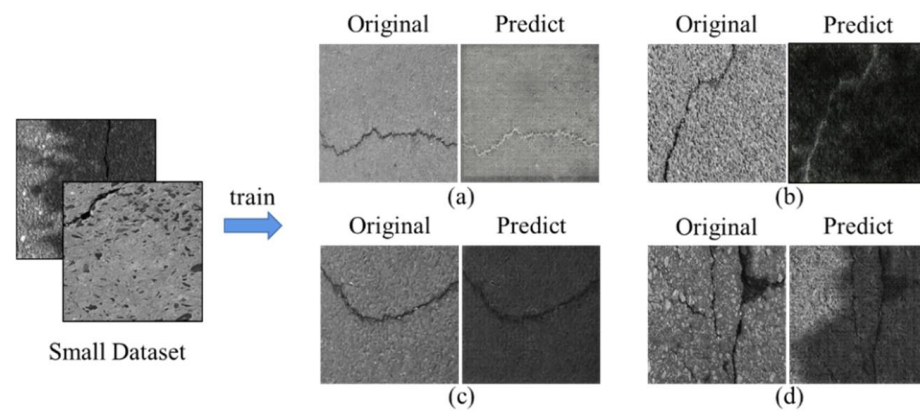
Figure 3. Problems arising from the training of small datasets in CycleGAN: (**a**) the loss of crack color information, (**b**) the loss of basic image information, (**c**) the overfitting caused by the small dataset, and (**d**) the loss of crack structure information.

The goal of the shadow-generation network is to give the mappings learned from the training samples *X* and *Y*, which include the mappings from *X* to *Y* for the shadow image generator *G* and from *Y* to *X* for the shadow-free image generator *F*. In addition, the network includes two discriminators, D_X and D_Y, which are mainly used to distinguish the authenticity of *X* and *Y* with their corresponding generated images. In this study, since the adopted dataset is a non-paired pavement dataset, our shadow mask information could not be generated from the original *X* domain and *Y* domain images, so we improved Equation (1) and separated our shadow information image from the generated image. The shadow semantic information *M* is expressed as follows:

$$M = \hat{Y} - X = Y - \hat{X} \tag{2}$$

where $\hat{Y}$ and $\hat{X}$ are the corresponding forgery image domains of *Y* and *X* generated by the two generators *G* and *F*, respectively.

According to the task requirements, as shown in Figure 4, we output the difference information between the input image and the generated image of the generator. Then we performed simple thresholding and normalization to obtain the shadow texture and display the shadow position and shape information in the form of a binary image.
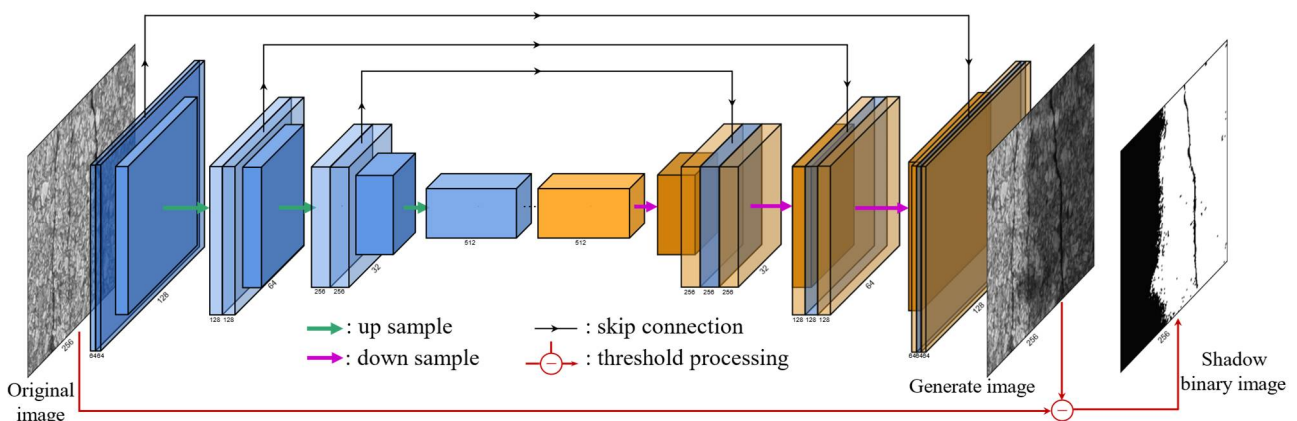

Figure 4. Illustration of generator network structure.

The CycleGAN in this paper mainly includes the following losses: adversarial loss, cycle consistent loss, identity loss, and the proposed texture self-supervised loss.

2.3.1. Adversarial Loss

The generative adversarial loss in CycleGAN adopts the same strategy as GAN, i.e., the mutual game between the generator and the discriminator is used for training. The generator tries to produce images that are more similar to the target domain image and tries to confuse the discriminator, which aims to classify the real picture more precisely from the generated one. Thus, the adversarial loss function is expressed as follows:

$$
\begin{aligned}
\mathcal{L}_{GAN}(G, D_Y, X, Y) \quad &= \mathbb{E}_{y \sim \mathcal{P}_{date}(y)}[\log D_Y(y)] \\
&+ \mathbb{E}_{x \sim \mathcal{P}_{date}(x)}[\log(1 - D_Y(G(x)))] \,,
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
\mathcal{L}_{GAN}(F, D_X, Y, X) \quad &= \mathbb{E}_{x \sim \mathcal{P}_{date}(x)}[\log D_X(x)] \\
&+ \mathbb{E}_{y \sim \mathcal{P}_{date}(y)}[\log(1 - D_X(F(y)))] \,,
\end{aligned}
\tag{4}
$$

where $\mathcal{L}_{GAN}(G, D_Y, X, Y)$ is the adversarial loss of generator G and discriminator D_Y, and similarly, the adversarial loss of generator F and discriminator D_X is $\mathcal{L}_{GAN}(F, D_X, Y, X)$. These two pairs of losses constitute the adversarial loss of CycleGAN.

2.3.2. Cycle Consistent Loss

Cycle consistent loss is an important loss of CycleGAN and also the main structure of its network. It ensures that the unpaired dataset has a relatively stable effect in training. The idea of circular consistency loss lies in the "translation" of images. When we carry out image semantic conversion, not only do we need to ensure that the two image domains can be transformed into each other, but also the converted forged domain image can be converted back to the original image domain through the corresponding generator, which is also the reason why it is named cycle consistent loss. The specific form of loss is as follows:

$$
\begin{aligned}
\mathcal{L}_{cyc}(G, F) \quad &= \mathbb{E}_{x \sim \mathcal{P}_{date}(x)}[\|F(G(x)) - x\|_1] \\
&+ \mathbb{E}_{y \sim \mathcal{P}_{date}(y)}[\|G(F(y)) - y\|_1] \,.
\end{aligned}
\tag{5}
$$

where $\mathcal{L}_{cyc}(G, F)$ is the cycle consistent loss, $F(G(X))$ is the reconstructed images from $G(X)$ by generator F, and, similarly, $G(F(y))$ is the reconstructed images from $F(y)$ by generator G.

2.3.3. Identity Loss

Identity loss is a loss function proposed in CycleGAN. Its function is to ensure the consistency of the color composition of the output image and the input image. To retain the color information of the input pavement image, this loss needs to be added to the task to reduce the color information loss of the generated image. This loss is expressed as follows:

$$
\begin{aligned}
\mathcal{L}_{identity}(G, F) \quad &= \mathbb{E}_{y \sim \mathcal{P}_{date}(y)}[\|G(y) - y\|_1] \\
&+ \mathbb{E}_{x \sim \mathcal{P}_{date}(x)}[\|F(x) - x\|_1].
\end{aligned}
\tag{6}
$$

where $\mathcal{L}_{identity}(G, F)$ is the identity loss, the image in the target domain of the generator is input into the generator, and the input image and the output image are compared to constrain and make it close to the identity map. In the formula, x is the target image of generator F, and y is the target image of generator G.

2.3.4. Texture Self-Supervised Loss

Although CycleGAN can be used for generating unpaired images, it still has certain limitations in the case of special tasks. For the shadow-generation task of the pavement crack image in this paper, when using CycleGAN to solve the mapping relationship in the image domain, because the image information is not symmetrical, the shadow generation task is closer to semantic generation than the semantic transformation task. However, CycleGAN lacks corresponding constraints on the shadow-generation task in the unpaired image domain, which often leads to the destruction of image color and structure. In this

case, it is usually improved by adding supervision. However, compared with adding labels to constrain the shape and position of generated information, the proposed method uses self-supervision of image-generated information, which can effectively reduce the task complexity caused by label production and make the loss setting more reasonable.

To achieve a better generation effect, this paper adds a new loss corresponding to the task based on the original CycleGAN loss function, which is shown as follows:

$$\mathcal{L}_{tss}(G,F) = \mathbb{E}_{x\sim\mathcal{P}_{date}(x),y\sim\mathcal{P}_{date}(y)}[\|M(x) - M(y)\|_1], \tag{7}$$

where $M(x)$ and $M(y)$ are binary images of shadow information output from generator G and generator F, respectively. This texture self-supervised loss, which is extracted from the cycle structure, constricts the generation and elimination of shadows by the two generators so that the generator can generate or eliminate shadow information more effectively.

Since this loss function takes effect during the mutual transformation of the two image domains, it makes one generator gradually converge in the process of solving the image mapping relationship, and the other generator must also converge in the same direction. This relationship is mutual. Thus, this set of losses makes the generator converge quickly. To adapt to other losses, it is necessary to set the weight of each loss according to the task requirements in order to optimize the overall training effect of the network.

2.3.5. Overall Loss Function

According to the constraint effect of each loss, this paper sets different weight parameters for the above losses to balance the network training effect. The overall loss function is expressed as follows:

$$\begin{aligned} L(G,F,D_X,D_Y) \quad &= \mathcal{L}_{GAN}(G,D_Y,X,Y) + \mathcal{L}_{GAN}(F,D_X,Y,X) \\ &+ \lambda_1 \mathcal{L}_{cyc}(G,F) \\ &+ \lambda_2 \mathcal{L}_{identity}(G,F) \\ &+ \lambda_3 \mathcal{L}_{tss}(G,F). \end{aligned} \tag{8}$$

The optimization objective of the network is as follows:

$$G^*, F^* = arg \min_{G,F} \max_{D_X,D_Y} \mathcal{L}(G,F,D_X,D_Y). \tag{9}$$

Through continuous training, find the best generation G and F and the most discriminative D_X and D_Y, which is the goal to optimize the network.

On the basis of the actual adjustment, we found that the training results had a certain difference when setting different weights, finally, the weight parameters in this paper were set as $\lambda_1 = 10$, $\lambda_2 = \lambda_3 = 5$, according to the training results. With this weight adjustment, the network was able to complete the shadow-generation task well.

3. Experiment

To verify the effect of the proposed texture self-supervised CycleGAN (CycleGAN-TSS) in the pavement shadow generation task, we collated and manually created a small number of corresponding shadow image datasets to study the shadow-generation effect of the network under the training of small datasets and compare the generation effect of the original CycleGAN under the same situation. Finally, to study the effect of dataset expansion on crack recognition, we used the shadow pavement crack image generated by the network to expand the pavement image dataset. The crack-segmentation effect of the U-Net model on shadow pavement pictures before and after dataset expansion was studied to determine the improvement of segmentation accuracy by crack-dataset expansion under shadow interference.

3.1. Experimental Environment and Data Sources

The experiments were conducted on a PC equipped with an Intel(R) Core i9-10900k CPU and an NVIDIA 3090, 24G GPU. Software environment: Windows 10, CUDA 11.1, CUDNN-v8.0.4, TensorFlow-GPU2.4, and Python 3.8.

Given the lack of image datasets with pavement shadows and cracks, to verify the stability of the shadow generation of the proposed CycleGAN-TSS in the case of pavement images with cracks, in this study, a pavement-crack-image dataset with the shadow was constructed by using three public datasets, namely CRACK500, CFD, and CrackTree200, as the main data sources. The specific production method of shadowed images is described in Section 2.1. Eighty shadow images were selected from each public dataset for model training, respectively.

As regards the dataset used to train U-Net and verify the augmentation effect, the pavement images with shadows were all generated by CycleGAN-TSS on the CRACK500 dataset.

3.2. Shadow Generation Network Training and Results

Following the training method of the original CycleGAN, we set the batch size of the network image input to 4, used the Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.99$ for parameter update, and set the initial learning rate to 2×10^{-4}. As shown in Table 2, the corresponding training images of the three datasets were fed into the network.

Table 2. Training set inputs for generating networks.

Dataset	Original Image	Composite Image
CRACK500	80	80
CFD	80	46
cracktree200	80	80

For ease of illustration, the original information of the pavement crack image was understood as the basic content information of the image, and the shadow information generated by the generative network was understood as the texture information. In CycleGAN-TSS, texture information was also output as shadow binary images for attention.

Figure 5 shows the images with shadow texture information generated by the forward mapping of the proposed generative network on the original image, and the shadow texture information in the binary image corresponds to the shape and position information of the shadow of the generated image. We found that the proposed model gradually generates the pavement shadow images closer to the real images, with the number of training epochs reaching 50. It can also be seen from the binary images that, with the progress of training epochs, the position and shape information of shadows were more similar to the distribution of the real shadowed images of the pavement surface. However, as the number of training epochs is 70, the effect of generated shadow images is not better than that of 50. Therefore, we can confirm that the model generation effect reached convergence when the training times reached 50 epochs.

To further verify the effectiveness of CycleGAN-TSS, the original CycleGAN and CycleGAN-TSS were trained on small unpaired datasets with the same training parameters and the same number of training epochs. By comparing the shadow-generation effects of the two networks, it was intuitively shown that CycleGAN-TSS can produce a better result on the target task.
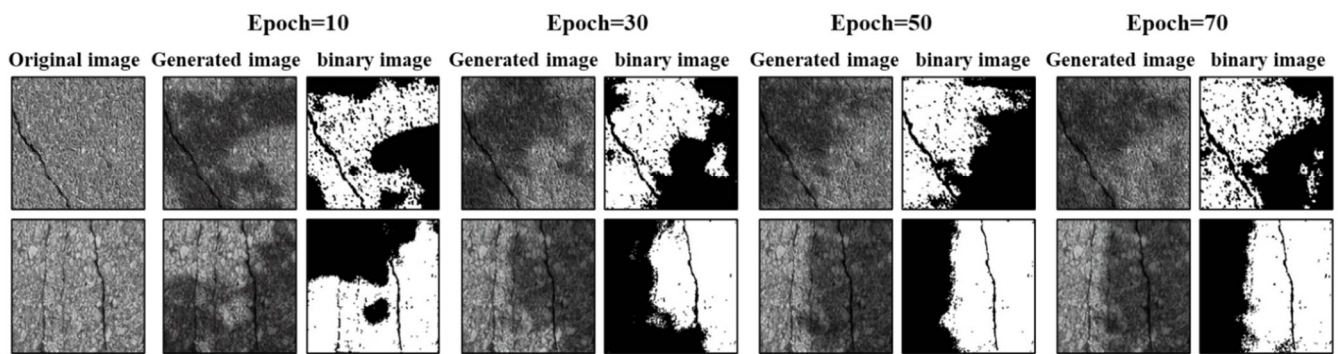
Figure 5. Examples of original images, generated images, and shadow texture information under different training epochs.

As shown in Figure 6, it was found through comparison that, when the shadowed image was generated with cracks, the original CycleGAN confused the color information of shadows and cracks during training, thus leading to color and structural damage in the generated picture. With the deepening of the training degree, the original CycleGAN also produced a certain training effect. However, when the generator and discriminator gradually converged, the problems of image noise and crack color loss were not solved, especially when the crack-interference information in the image was more complex. By comparing the image-generation results of CycleGAN-TSS with the image-generation effect of the original CycleGAN, the following conclusions can easily be drawn:

- With the same training conditions, compared with the original CycleGAN, CycleGAN-TSS has a faster convergence speed. The reason is that it adds a self-supervised mechanism to introduce the prior condition of image-generation effect convergence into the network, i.e., the texture generation cycle consistent condition. In other words, the shadow texture information generated and removed by the shadow-generation network should be consistent at least in position and shape;
- In terms of the final image-generation effect, CycleGAN-TSS produces less image noise outside the shadow texture region of the generated image. The reason is that the texture changes of the two processes of shadow generation and shadow removal are constrained by a self-supervised loss, which limits the changes of the image to the same region for both shadow generation and shadow removal. Combined with the participation of adversarial loss, the discriminator restricts the authenticity of shadow generation, which enables it to achieve a better shadow-generation effect, while also preserving more complete content information of non-shadow regions.

For semantic-generation tasks, especially those similar to shadow-generation tasks, such tasks themselves pay more attention to the position and shape information of the generated semantics, thus causing, the new CycleGAN with a texture self-supervised loss function to have better generation effects in such tasks.

In Figure 7a,b show the shadow-generation effects of two datasets, namely Crack-Tree200 and CFD, trained with CycleGAN-TSS, respectively. Similar to the CRACK500 dataset, these two datasets used only 80 synthetic images of pavement cracks with shadows for training, and the task of shadow generation was also completed well in the case of this small dataset.

In general, CycleGAN-TSS applied to the task of generating shadows from pavement crack images can produce better results than the original CycleGAN. Therefore, we applied the network to the augmentation task of shadow images in an attempt to improve the accuracy of the pavement crack segmentation task in the case of shadow interference. As a result of the experimental setup of the segmentation task in this paper, the CRACK500 dataset with the most sufficient image data was selected for shadow image augmentation, and the augmented dataset was fed into the U-Net network for training to study the accuracy improvement brought about by our proposed image data augmentation method.
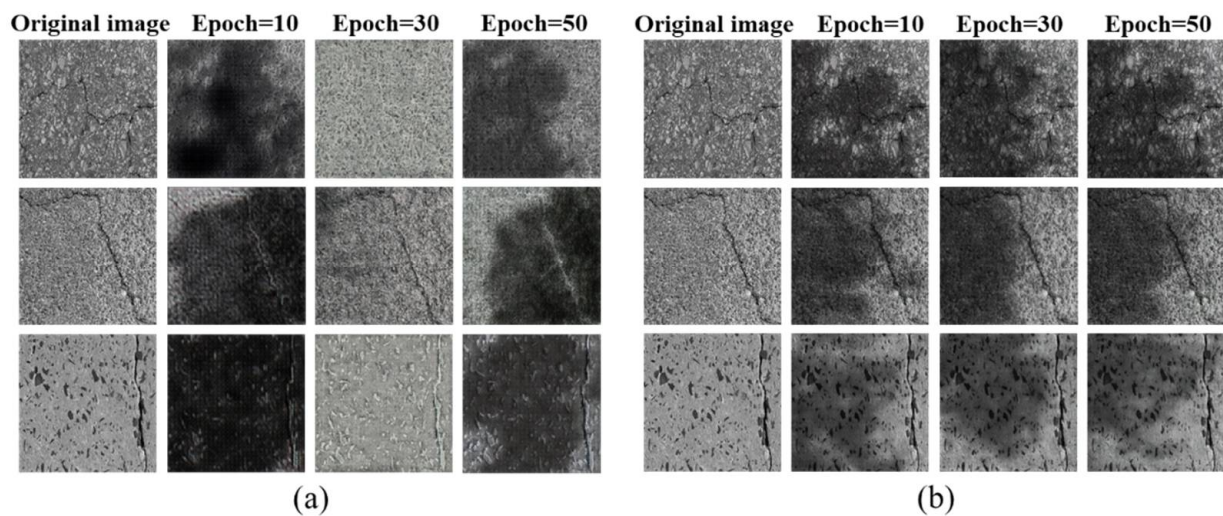
Figure 6. Comparison of original CycleGAN- and CycleGAN-TSS-generated images: (**a**) from left to right, we have the original images and the images generated by the original CycleGAN with training epochs of 10, 30, and 50, respectively; (**b**) from left to right, we have the original images and the images generated by CycleGAN-TSS with training epochs of 10, 30, and 50, respectively.
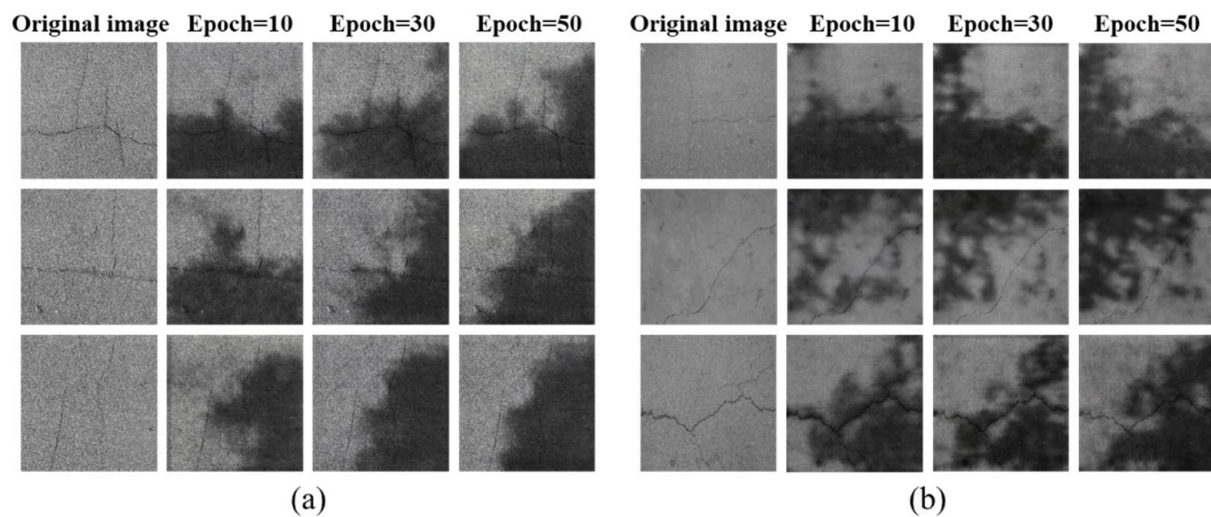

Figure 7. Generation effect of two datasets of the CycleGAN-TSS model: (**a**) CrackTree200 original images and generated images at epoch = 10, 30, and 50; (**b**) CFD original images and generated images at epoch = 10, 30, and 50.

3.3. Segmentation Network Validation Results

To study the influence of the augmentation of the pavement crack dataset with shadows on the anti-shadow interference effect of the segmentation network, we conducted crack segmentation and recognition experiments on the original pavement image dataset and the corresponding image dataset with shadows generated by CycleGAN-TSS to analyze the actual effect of the pavement shadow image augmentation.

In this study, four datasets were assigned to the training task: T1—a training dataset containing 500 original images and a test set containing 200 original images; T2—a training dataset containing 500 original images and a test set containing 200 images with a 1:1 mixture of original and shadowed images; T3—a training dataset containing 350 original images and 150 shadowed images and a test set containing 200 images with a 1:1 mixture of original and shadowed images; T4—a training dataset with 350 original images and 150 shadowed images and a test set with 200 original images, as shown in Table 3.

Table 3. Composition of the four datasets for training and testing.

Dataset	Train Set		Test Set	
	Original Image	Augmented Image	Original Image	Augmented Image
T1	500	0	200	0
T2	500	0	100	100
T3	350	150	100	100
T4	350	150	200	0

To reduce the interference of other factors, the 500 original images used for training and 350 original images in the mixture set were the same, and the remaining 150 original images from the 500 were sent to CycleGAN-TSS to generate 150 pavement images with shadows in the mixture set. The mixed image test set was also generated using this strategy.

The above datasets were fed into U-Net for training, whereby the T1 and T2 datasets had the same training set and we trained it only once; the same is true for the T3 and T4 datasets. The Acc and Loss evaluation metrics used for training and MIoU changes during the training process are shown in Figure 8. According to Figure 8, it can be seen that the model reached convergence when the number of training times was 200. The final model obtained by training was used to segment the cracks of the test-set images, and the final segmentation effect of the test set is displayed by the main evaluation index.
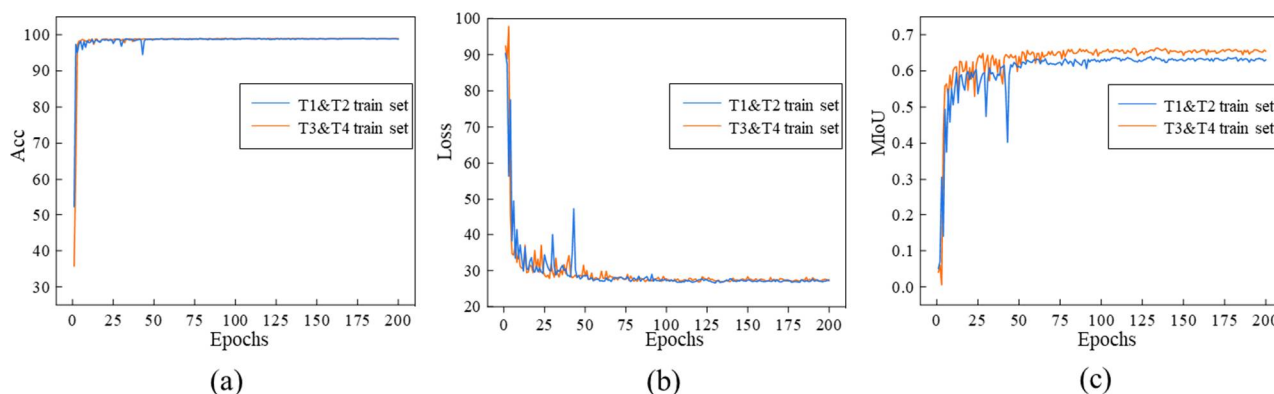

Figure 8. U-Net network training process: (**a**) the change in Acc with training epochs, (**b**) the change in Loss with training epochs, and (**c**) the change in MIoU with training epochs.

Figure 9 shows that U-Net was able obtain good segmentation results in the crack segmentation of shadow-free images under the training of the T1 dataset. Under the training of the T2 dataset, the same segmentation network model as T1 was obtained, but the segmentation accuracy was significantly reduced when the crack was segmented under the test shadow condition. The segmentation results of T3 and T4 show that, when the dataset with shadow image augmentation was involved in the training, the trained model achieved good segmentation results for both crack detection with shadow and crack detection without shadow.

The above illustrates that, when the dataset participating in training lacked shadow images, the dataset sent to the network for training was unbalanced, thus making it difficult for the network to detect and segment pavement crack images with shadows. However, if we augmented the dataset by generating shadows from part of the original images through CycleGAN-TSS, we were able to effectively improve the segmentation effect of the U-Net network under the same training conditions.
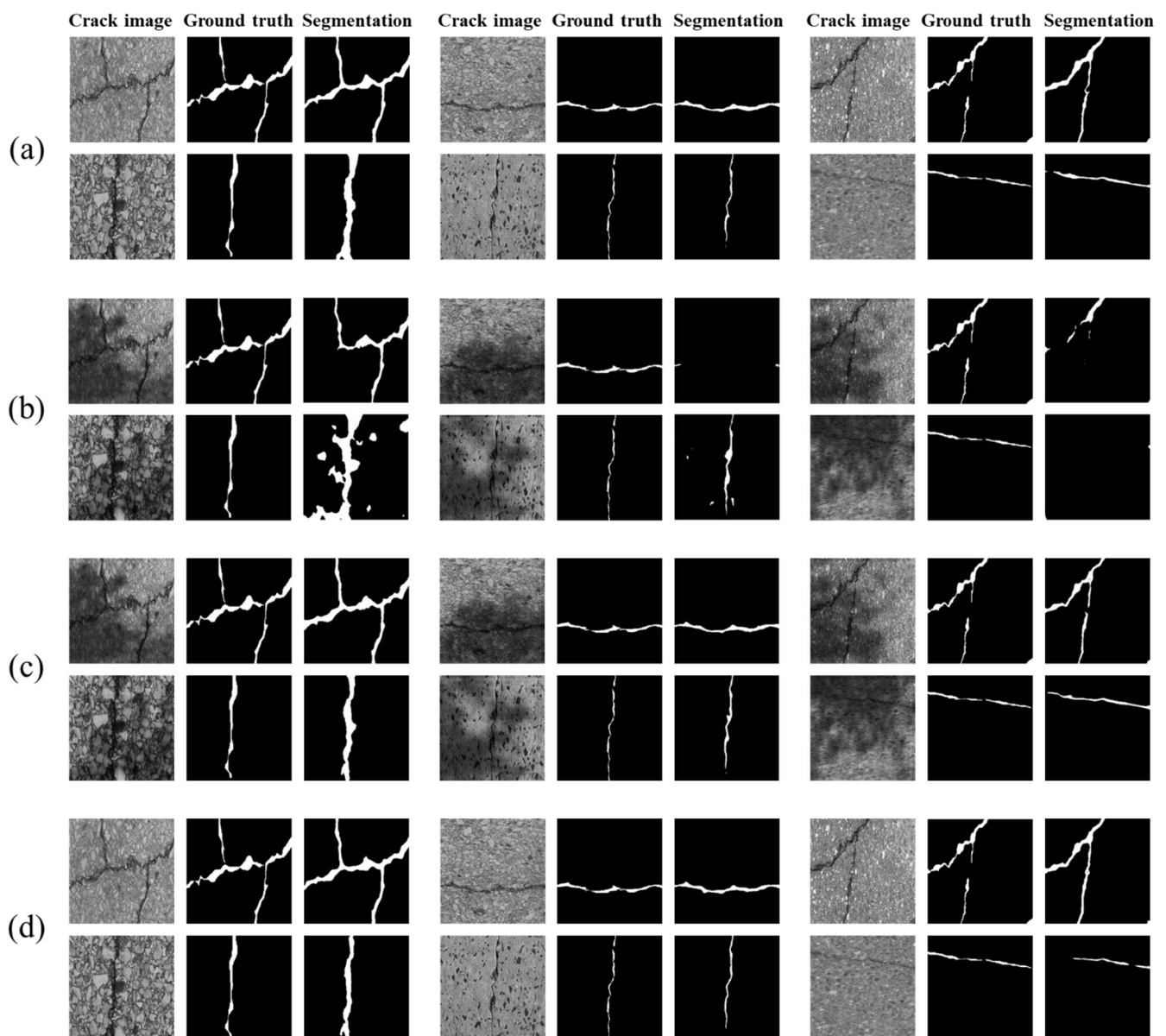
Figure 9. Segmentation effect before and after data augmentation: (**a**) the training segmentation effect of the T1 dataset, (**b**) the training segmentation effect of the T2 dataset, (**c**) the training segmentation effect of the T3 dataset, and (**d**) the training segmentation effect of the T4 dataset.

To evaluate the effect of the augmented shadow dataset more accurately, we introduced the following general segmentation effect evaluation indicators: Acc (accuracy), Recall, F1-score, and MIoU (Mean Intersection over Union).

To briefly explain the above evaluation metrics, we first defined the number of pixels correctly classified as cracks as *TP* (true positive), the number of pixels incorrectly classified as cracks as *FP* (false positive), and the number of pixels correctly classified as background as *TN* (true negative). If the number of pixels misclassified as the background is *FN* (false negative), then the above evaluation metric can be expressed as follows:

$$
\begin{aligned}
Acc &= \frac{TP+TN}{TP+FP+TN+FN}, \\
Recall &= \frac{TP}{TP+FN}, \\
F1 &= 2\frac{Precision \times Recall}{Precision + Recall}, \\
MIOU &= \frac{TP}{TP+FP+FN},
\end{aligned}
\tag{10}
$$

where the precision is denoted as $Precision = TP/(TP + FP)$; thus, our testing results on these four datasets are shown in Table 4.

Table 4. Effect of crack segmentation in the test set with different training datasets.

Dataset	Acc (%)	Recall	F1-Score	MIoU
T1	98.25	78.46	77.39	63.12
T2	97.16	64.15	63.29	46.29
T3	98.16	77.11	76.18	61.53
T4	98.30	79.78	77.69	63.52

According to Table 4, it can be found that the model trained under the dataset of the original image was able to achieve a better crack segmentation effect for the test set of the pavement crack images without shadows. However, under this model, the three main indicators, namely Recall, F1-score, and MIoU, were all reduced to different degrees when segmenting the dataset with shadow images. It was observed that the training model invested in part of the augmented dataset exhibited better segmentation accuracy for the pavement crack images with shadows than the training model of the original image dataset. Moreover, the model still exhibited good accuracy for the segmentation of the test set of non-shadow images. In summary, we were able to draw the following conclusions:

- Shadows have a great influence on the segmentation of pavement cracks, mainly because the structure and color information of cracks is easily confused with shadows;
- By augmenting the shadow dataset, the ability of the network to resist shadow interference can be effectively improved, and the crack segmentation effect of the non-shadow pavement images can be well maintained.

For the pavement-shadow-image augmentation task, the CycleGAN-TSS in this paper exhibits a good performance and can stably generate shadow-interference images with appropriate effects. Moreover, the generated pavement shadow images also improve the anti-shadow interference ability of pavement crack segmentation.

4. Conclusions

In this paper, we proposed a new method for shadowed-pavement-image-data augmentation and proposed a texture self-supervised structure based on CycleGAN to improve the generation effect of shadow texture:

1. Texture self-supervised CycleGAN (CycleGAN-TSS) introduces the prior knowledge of generator convergence, reduces the problem of poor generation effect caused by the similarity between the semantic information of cracks and the semantic information of shadows, and improves the stability of shadow images generated by the network;

2. In this study, we conducted experiments on three different datasets to verify the effect of the generated network. Compared with the original CycleGAN, the improved CycleGAN-TSS pays more attention to the shape and location information of the cycle generation semantics, thus making our network better able to generate pavement shadow images in the case of small datasets;

3. In summary, the proposed CycleGAN-TSS is more suitable for the data augmentation of pavement shadow images in real-world conditions. We used the proposed method to augment pavement shadow images and fed them into the crack segmentation network U-Net for training. The results show that the augmented method can effectively improve the recognition accuracy of crack detection under shadow interference.

Aside from improving the effect of automatic pavement detection in real-world conditions, the method proposed in this paper can also be used to solve similar problems and provide a possible path for solving pavement-detection tasks in complex situations in the future. In addition, CycleGAN-TSS will also have a certain effect on other generation tasks that weakly focus on color information. Therefore, we will continue to expand the application of this network in other industrial dataset augmentation tasks.

References

1. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 30 June 2016; pp. 770–778.
2. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]
3. Chen, Y.; Bai, Y.; Zhang, W.; Mei, T. Destruction and Construction Learning for Fine-Grained Image Recognition. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 5152–5161. [CrossRef]
4. Li, X.; Zhao, L.; Wei, L.; Yang, M.-H.; Wu, F.; Zhuang, Y.; Ling, H.; Wang, J. DeepSaliency: Multi-Task Deep Neural Network Model for Salient Object Detection. *IEEE Trans. Image Process.* **2016**, *25*, 3919–3930. [CrossRef] [PubMed]
5. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
6. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
7. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef] [PubMed]
8. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241. [CrossRef]
9. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
10. Rao, Z.; Tung, P.-Y.; Xie, R.; Wei, Y.; Zhang, H.; Ferrari, A.; Klaver, T.P.C.; Körmann, F.; Sukumar, P.T.; Kwiatkowski da Silva, A.; et al. Machine learning–enabled high-entropy alloy discovery. *Science* **2022**, *378*, 78–85. [CrossRef]
11. Fan, W.; Chen, Y.; Li, J.; Sun, Y.; Feng, J.; Hassanin, H.; Sareh, P. Machine learning applied to the design and inspection of reinforced concrete bridges: Resilient methods and emerging applications. *Structures* **2021**, *33*, 3954–3963. [CrossRef]
12. Zhang, P.; Fan, W.; Chen, Y.; Feng, J.; Sareh, P. Structural symmetry recognition in planar structures using Convolutional Neural Networks. *Eng. Struct.* **2022**, *260*, 114227. [CrossRef]
13. Gopalakrishnan, K. Deep Learning in Data-Driven Pavement Image Analysis and Automated Distress Detection: A Review. *Data* **2018**, *3*, 28. [CrossRef]
14. Nhat-Duc, H.; Nguyen, Q.-L.; Tran, V.-D. Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network. *Autom. Constr.* **2018**, *94*, 203–213. [CrossRef]
15. Ji, A.; Xue, X.; Wang, Y.; Luo, X.; Xue, W. An integrated approach to automatic pixel-level crack detection and quantification of asphalt pavement. *Autom. Constr.* **2020**, *114*, 103176. [CrossRef]
16. Li, P.; Xia, H.; Zhou, B.; Yan, F.; Guo, R. A Method to Improve the Accuracy of Pavement Crack Identification by Combining a Semantic Segmentation and Edge Detection Model. *Appl. Sci.* **2022**, *12*, 4714. [CrossRef]
17. Wang, Y.; Zhang, J.Y.; Liu, J.X.; Zhang, Y.; Chen, Z.P.; Li, C.G.; He, K.; Yan, R.B. Research on Crack Detection Algorithm of the Concrete Bridge Based on Image Processing. *Procedia Comput. Sci.* **2019**, *154*, 610–616. [CrossRef]
18. Qiang, S.; Guoying, L.; Jingqi, M.; Hongmei, Z. An edge-detection method based on adaptive canny algorithm and iterative segmentation threshold. In Proceedings of the 2016 2nd International Conference on Control Science and Systems Engineering (ICCSSE), Singapore, 27–29 July 2016; IEEE: Singapore, 2016; pp. 64–67. [CrossRef]
19. 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]
20. Wang, W.; Zhang, X.; Hong, H. Pavement Crack Detection Combining Non-Negative Feature with Fast LoG in Complex Scene. In Proceedings of the Ninth International Symposium on Multispectral Image Processing and Pattern Recognition (MIPPR2015), Enshi, China, 31 October–1 November 2015; p. 98120L. [CrossRef]

21. Huyan, J.; Li, W.; Tighe, S.; Deng, R.; Yan, S. Illumination Compensation Model with *k*-Means Algorithm for Detection of Pavement Surface Cracks with Shadow. *J. Comput. Civ. Eng.* **2020**, *34*, 04019049. [CrossRef]
22. Liu, Z.; Yin, H.; Mi, Y.; Pu, M.; Wang, S. Shadow Removal by a Lightness-Guided Network With Training on Unpaired Data. *IEEE Trans. Image Process.* **2021**, *30*, 1853–1865. [CrossRef] [PubMed]
23. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
24. Nanni, L.; Paci, M.; Brahnam, S.; Lumini, A. Feature transforms for image data augmentation. *arXiv* **2014**, arXiv:2201.09700. [CrossRef]
25. Antoniou, A.; Storkey, A.; Edwards, H. Data Augmentation Generative Adversarial Networks. *arXiv* **2018**, arXiv:1711.04340.
26. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *arXiv* **2014**, arXiv:1406.2661. [CrossRef]
27. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. *arXiv* **2018**, arXiv:1611.07004.
28. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2242–2251. [CrossRef]
29. Sandfort, V.; Yan, K.; Pickhardt, P.J.; Summers, R.M. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci. Rep.* **2019**, *9*, 16884. [CrossRef] [PubMed]
30. Niu, S.; Li, B.; Wang, X.; He, S.; Peng, Y. Defect attention template generation cycleGAN for weakly supervised surface defect segmentation. *Pattern Recognit.* **2022**, *123*, 108396. [CrossRef]
31. Choi, W.; Heo, J.; Ahn, C. Development of Road Surface Detection Algorithm Using CycleGAN-Augmented Dataset. *Sensors* **2021**, *21*, 7769. [CrossRef] [PubMed]
32. Wang, L.; Wang, L.; Chen, S. ESA-CycleGAN: Edge feature and self-attention based cycle-consistent generative adversarial network for style transfer. *IET Image Process.* **2022**, *16*, 176–190. [CrossRef]
33. Xu, Z.; Qi, C.; Xu, G. Semi-Supervised Attention-Guided CycleGAN for Data Augmentation on Medical Images. In Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, 18–21 November 2019; pp. 563–568. [CrossRef]
34. Jiangsha, A.; Tian, L.; Bai, L.; Zhang, J. Data augmentation by a CycleGAN-based extra-supervised model for nondestructive testing. *Meas. Sci. Technol.* **2022**, *33*, 045017. [CrossRef]
35. 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]
36. Yang, F.; Zhang, L.; Yu, S.; Prokhorov, D.; Mei, X.; Ling, H. Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 1525–1535. [CrossRef]
37. Shi, Y.; Cui, L.; Qi, Z.; Meng, F.; Chen, Z. Automatic Road Crack Detection Using Random Structured Forests. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 3434–3445. [CrossRef]