

Article

Sample Expansion and Classification Model of Maize Leaf Diseases Based on the Self-Attention CycleGAN

Hongliang Guo ¹, Mingyang Li ¹, Ruizheng Hou ¹, Hanbo Liu ¹, Xudan Zhou ², Chunli Zhao ², Xiao Chen ^{1,*} and Lianxing Gao ^{3,*}

¹ College of Information Technology, Jilin Agricultural University, Changchun 130118, China; guohongliang@jlau.edu.cn (H.G.); lim959226@gmail.com (M.L.); hrz13578561921@163.com (R.H.); liuh46807@gmail.com (H.L.)

² College of Forestry and Grassland, Jilin Agricultural University, Changchun 130118, China; zhouxudan@jlau.edu.cn (X.Z.); zhaochunli@jlau.edu.cn (C.Z.)

³ College of Engineering and Technology, Jilin Agricultural University, Changchun 130118, China

* Correspondence: chenxiao@jlau.edu.cn (X.C.); lianxing_gao@126.com (L.G.)

Abstract: In order to address the limited scale and insufficient diversity of research datasets for maize leaf diseases, this study proposes a maize disease image generation algorithm based on the cycle generative adversarial network (CycleGAN). With the disease image transfer method, healthy maize images can be transformed into diseased crop images. To improve the accuracy of the generated data, the category activation mapping attention mechanism is integrated into the original CycleGAN generator and discriminator, and a feature recombination loss function is constructed in the discriminator. In addition, the minimum absolute error is used to calculate the differences between the hidden layer feature representations, and backpropagation is employed to enhance the contour information of the generated images. To demonstrate the effectiveness of this method, the improved CycleGAN algorithm is used to transform healthy maize leaf images. Evaluation metrics, such as peak signal-to-noise ratio (PSNR), structural similarity (SSIM), Fréchet inception distance (FID), and grayscale histogram can prove that the obtained maize leaf disease images perform better in terms of background and detail preservation. Furthermore, using this method, the original CycleGAN method, and the Pix2Pix method, the dataset is expanded, and a recognition network is used to perform classification tasks on different datasets. The dataset generated by this method achieves the best performance in the classification tasks, with an average accuracy rate of over 91%. These experiments indicate the feasibility of this model in generating high-quality maize disease leaf images. It not only addresses the limitation of existing maize disease datasets but also improves the accuracy of maize disease recognition in small-sample maize leaf disease classification tasks.

Keywords: cycle-consistent adversarial networks; attention mechanism; maize leaf disease identification; feature recombination; computer vision



Citation: Guo, H.; Li, M.; Hou, R.; Liu, H.; Zhou, X.; Zhao, C.; Chen, X.; Gao, L. Sample Expansion and Classification Model of Maize Leaf Diseases Based on the Self-Attention CycleGAN. *Sustainability* **2023**, *15*, 13420. <https://doi.org/10.3390/su151813420>

Academic Editor: Hossein Azadi

Received: 28 July 2023

Revised: 3 September 2023

Accepted: 4 September 2023

Published: 7 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Crop diseases pose a significant threat to the quality and quantity of global agricultural production, leading to a substantial reduction in the economic productivity of crops. This presents a major challenge to food security, with catastrophic crop diseases exacerbating the existing global food shortage. Additionally, agriculture serves as a source of raw materials for textiles, chemical products, and pharmaceuticals. From the 1960s to the early 21st century, the amount of land utilized for agriculture increased by only 10%, and agricultural production grew three-fold. Looking ahead, considering the limited availability of land for agricultural use, the solution to food insecurity lies in enhancing the productivity of existing farmlands. This will require the cultivation of high-yielding, faster-maturing, drought-resistant, and disease-tolerant crop varieties. It is anticipated that these developments will

result in a decline in the use of chemical substances in agriculture, replaced instead by an emphasis on early and accurate detection of crop diseases and pests.

Maize, as a crucial component of our country's grain cultivation, boasts the largest planting area and highest overall yield. It serves as a vital source for both animal feed and industrial raw materials, playing a significant strategic role in ensuring our national food security. Consequently, the prevention and control of maize diseases are of paramount importance. These diseases primarily affect the leaves of the corn plants. Attempting to observe these diseases with the naked eye across vast planting areas or resorting to preventive measures based solely on previous cultivation experiences not only fails to provide a clear understanding of the disease situation but also results in a significant waste of resources on disease prevention efforts. However, diseases such as maize rust, gray leaf spot, and leaf spot have severely affected maize production [1]. Traditional disease diagnosis heavily relies on agricultural experts or technicians who evaluate diseases based on their expertise, leading to a time-consuming, labor-intensive, and inefficient process. This method struggles to meet the requirements of real-time and accurate control and prevention of diseases [2]. With the application of computer vision to agricultural disease images, technologies for artificial intelligence image recognition have provided significant assistance in the early detection of maize diseases [3]. In 2011, Kai S et al. [4] processed and analyzed images of corn diseases. They utilized the YCbCr color space technique and gray-level co-occurrence matrix in combination with texture features of corn disease to segment and extract features of lesions. The accuracy of disease classification reached up to 98% using a backpropagation neural network. In 2012, Kulkarni et al. [5] employed artificial neural networks (CNNs) and several image processing techniques to introduce a timely and precise method for plant disease detection. By leveraging Gabor filters for feature extraction and neural networks for classification, they attained a recognition rate of up to 91%.

Traditional methods for plant disease recognition are based on image processing and computer vision techniques, typically involving the extraction of features such as the shape, texture, and color of disease lesions. These methods heavily rely on the domain expertise within the field of agricultural diseases, resulting in relatively low recognition efficiency. In recent years, the rapid development of deep learning has attracted many researchers to conduct relevant studies aimed at improving the accuracy of plant disease identification. These technologies are based on deep learning with convolutional neural networks as their core. However, deep learning relies on data, and the size of the dataset significantly impacts the quality of the training outcomes [6]. Obtaining a sufficient amount of data is an essential requirement for successfully accomplishing tasks and creating a neural network model that is capable of learning possible distributions [7]. Currently, the scarcity of data remains a significant obstacle in the development of deep learning technology. Building the necessary dataset takes time to accumulate, and consequently, the inadequacy of training samples significantly impacts the accuracy of maize disease recognition [8].

The traditional data augmentation methods typically include undersampling, oversampling, and image transformations [9–11]. While they can—to some extent—adjust the inter-class distribution of samples, they fail to take into account the overall distribution characteristics of the samples. However, maize leaf diseases are often characterized by their color distribution and contrast. Currently, data augmentation through geometric transformations and cropping [12] can suffice for most recognition tasks. However, in practical applications, due to factors such as the small infection area of plant diseases and varying degrees of severity, conventional data augmentation methods often struggle to accurately capture information about disease regions. The generated images may not exhibit distinct disease features and could potentially even degrade the performance of plant disease recognition.

In recent years, there have been significant breakthroughs in computer vision research [13,14], enabled by the rise of deep learning. One notable advancement is the proposal of a variational autoencoder (VAE) by Kingma et al. [15] in 2014. VAE is a generative network structure based on variational Bayes (VB) inference, which estimates the distribution of samples to generate similar ones. However, due to its pixel-level supervision of images, the VAE is limited in its ability to capture global information. This limitation often results in the overall blurriness of generated images. Another significant development in the field came in 2014, when Goodfellow et al. [16] introduced generative adversarial net (GAN). GAN consists of two players, a generator and a discriminator. The generator aims to generate fake samples that cannot be distinguished from real ones, whereas the discriminator's task is to correctly differentiate between real and generated samples [17]. By engaging in a game between the two, the final generated data are made to be indistinguishable from reality [18]. This method has found broad applications [19], such as data augmentation [20,21], image style transfer [22,23], image super-resolution [24,25], and text-to-image generation. GAN adopts an unsupervised learning approach, automatically learning from the source data to produce astonishing results without the need for manual labeling of the dataset [26–29]. However, the current GAN game process [30] indirectly establishes a relationship with real data through the discriminator, which fails to utilize the prior knowledge about the composition of input data. This leads to instability during training, poor quality of generated images, and potential mode collapse problems [31]. In 2016, Isola et al. [32] presented Pix2Pix, a GAN-based framework for supervised image-to-image translation. Pix2Pix utilizes conditional GAN (CGAN) [33] to guide the image generation process using conditional information. The model employs U-Net [34] as the generator and PatchGAN as the discriminator, successfully transforming labeled paired data while maintaining image structure consistency. However, Pix2Pix's reliance on paired training data poses challenges for tasks like artistic style transfer and object conversion [35]. In 2017, Zhu et al. [36] introduced the cycle generative adversarial network (CycleGAN) by combining GAN with the concept of dual learning. CycleGAN utilizes two generators and two discriminators to achieve cyclic image transformation, preserving content information using cycle consistency loss. This method, benefiting from the joint use of adversarial networks and a cycle-consistent structure, has demonstrated improved performance in image processing. Furthermore, CycleGAN does not require one-to-one pairing of data during training, making it widely applicable in image translation, style transfer, image enhancement, and related problems. Zhang et al. [37] enhanced the feature extraction capability of the traditional CycleGAN in 2023 by incorporating a self-attention module and atrous convolution multi-scale feature fusion module. They introduced a perceptual loss function into the model's loss function to enhance the texture perception of generated images. In 2022, Lu et al. [38] addressed the dataset imbalance issue using CycleGAN and improved the network performance by adding an efficient channel attention module. Hu et al. [39] proposed an improved CycleGAN framework for translating shortwave infrared face images to visible light face images, effectively overcoming the image modal differences caused by varying spectral characteristics and improving image observability. Additionally, Li et al. [40] enhanced the CycleGAN loss function by incorporating strong edge structure similarity, leading to color correction and the enhancement of underwater images.

To enhance the feature extraction capability of disease images, we introduce the class activation map (CAM) attention mechanism proposed by Zhou et al. [41] into the generator and discriminator of the CycleGAN, respectively. The CAM is a technique used to interpret deep learning models. In image classification tasks, CAM can help understand which regions of an image the model is focusing on during prediction. By generating class activation maps, it is possible to visualize the image regions that contribute to the model making specific classification decisions. This is highly useful for explaining the model's decision-making process and enhancing its interpretability. Typically, CAM is generated by multiplying globally average-pooled feature maps with the model's weights, emphasizing image regions relevant to specific classes. The CAM attention mechanism can enhance the

feature extraction capability of specific regions. In addition, a feature recombination loss function is used in the discriminator to optimize the edge information of the generated maize disease leaf images, thus improving the quality of the generated images. This method significantly improves the accuracy of maize disease classification in the maize leaf disease identification task.

2. Materials and Methods

2.1. The Principle of CycleGAN

CycleGAN is a variant of the GAN network that utilizes two transformation networks, denoted as F and G , to facilitate data transformation between different domains. DiscoGAN [42] and dual GAN [43] employ similar ideas, which partially alleviate the demands placed on the data. Each transformation network is trained by a separate GAN network with its respective generator. The primary objective of the discriminator is to distinguish between the generated data and real data. The optimization objective function of CycleGAN is depicted in Equation (1):

$$L_{\text{CycleGAN}} = L_{\text{gan}}(G, D_Y, X, Y) + L_{\text{gan}}(F, D_X, X, Y) + \mu L_{\text{cyc}}(G, F) \quad (1)$$

where D_Y represents the discriminator from domain Y to domain X , and is used to distinguish the differences between images generated from domain Y and real images from domain X . D_X represents the discriminator from domain X to domain Y , and is used to distinguish the differences between images generated from domain X and real images from domain Y . X signifies the input image, which can be a real image from domain X . Y represents the target image, which can be a real image from domain Y . The purpose of this loss function is to prompt the generator G to transform input image X into target image Y , and then generator F transforms the target image Y back to the original image X . Cycle consistency loss ensures the consistency of the cycle transformation by comparing the differences between generated images and original images. μ is a weighting parameter used to balance the importance of cycle consistency loss and adversarial loss. Moreover, $L_{\text{gan}}(G, D_Y, X, Y)$ and $L_{\text{gan}}(F, D_X, X, Y)$ represent the losses for two generative adversarial networks, while $L_{\text{cyc}}(G, F)$ represents the loss for the cycle-consistent network.

2.2. Improvements to the CycleGAN Model

By incorporating attention mechanisms into the CycleGAN, an improved maize disease image generation model was constructed. Additionally, a feature recombination loss function was introduced to enhance the model's performance. The model comprises two generators and two discriminators. Generator G transforms healthy maize leaf images into diseased maize leaf images, while generator F operates in the reverse direction, converting diseased maize leaf images into healthy ones. The discriminators evaluate the authenticity of the input healthy maize images and diseased maize leaf images, respectively. The overall structure of the model is shown in Figure 1. During training, a cycle consistency loss function is incorporated to support the performance of generator G_s , ensuring that the network achieves a closed-loop training state. Attention mechanism modules are incorporated into generator G_s and discriminator D_s in Figure 1 to enhance the conversion of healthy maize leaf images into maize disease leaf images. The network architectures of G_s and D_s will be discussed in detail in Sections 2.2.2 and 2.2.3.

2.2.1. Introduction of Attention Mechanism

The attention mechanism employed in the generator and discriminator is the class activation map (CAM) soft attention method. This method calculates the spatial average values of individual neurons from the feature maps obtained through convolution and global average pooling. These average values are then linearly weighted to generate the required class activation map. The localization of CAM within the image is achieved by utilizing category information and weight information specific to the image. Upsampling CAM produces a feature map of the same size as the input image, determining the position

region relevant to the label category. The calculation of the CAM feature map adopts a global approach, encompassing the entire feature map. The generation process based on the CAM attention feature map is illustrated in Figure 2.

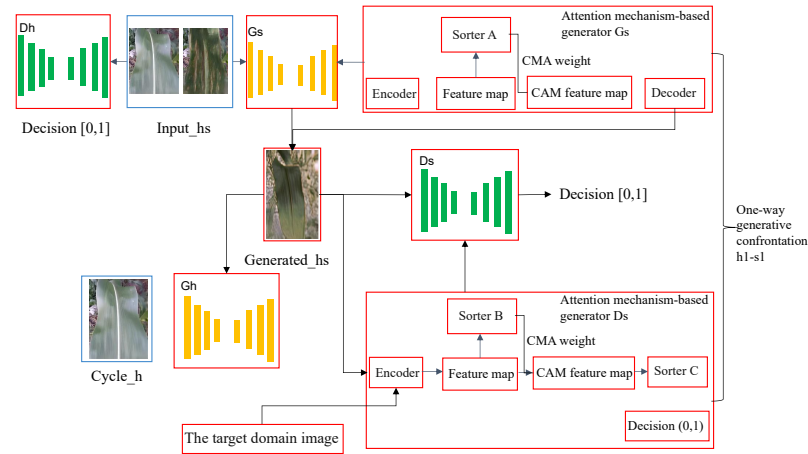


Figure 1. Improved CycleGAN model structure.

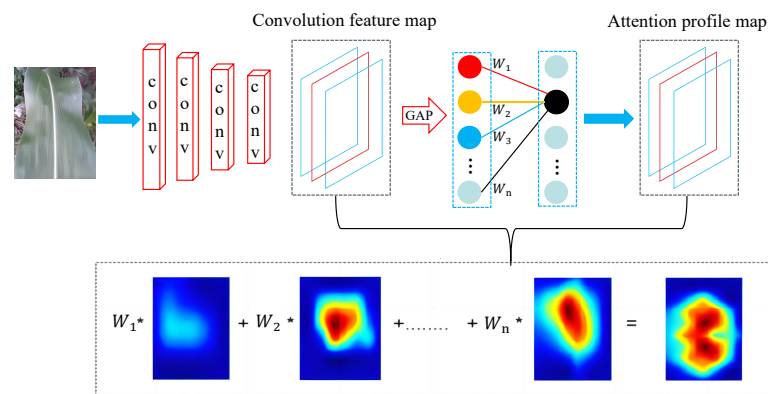


Figure 2. Generation process of the attention feature map based on CAM.

The CAM attention mechanism operates on convolutional feature maps obtained through convolution. Let $H(x_i)$ represent the feature map obtained after convolution, and $f_k(x, y)$ denote the activation of unit k in the last convolution layer at spatial location (x, y) . Subsequently, global average pooling is applied to compute the result F_k . This result is then classified using softmax to obtain the score S_c for a given category c . Finally, the region relevance P_c is calculated using the following formula:

$$P_c = \frac{\exp(S_c)}{\sum_c \exp(S_c)} \quad (2)$$

$$S_c = \sum_k w_k^c \sum_{x,y} f_k(x, y) \quad (3)$$

$$F_k = \sum_{x,y} f_k(x, y) \quad (4)$$

$$Q_c = \sum_{i=1}^N P_c H(x_i) \quad (5)$$

The attention on the region of category c is denoted by P_c , while the class activation heatmap obtained through global average pooling and softmax applied to the convolutional

feature map in the region of category c is represented by S . The weight information w_k^c corresponds to F_k , and the output of the CAM attention module is denoted by Q_c .

2.2.2. Generator Network Architecture

In the original CycleGAN model, the generator comprises an encoder, a transformer, and a decoder. The original CycleGAN utilized six ResBlock structures in its transformer. Upon integrating the CAM soft attention mechanism, the six ResBlock structures in the transformer were subdivided and added to the encoder and decoder of the generator, respectively. Furthermore, classifier A was included between the encoder and decoder of the generator. Figure 3a depicts the generator structure based on the attention mechanism. In the enhanced generator, as depicted in Figure 3a, the input images are sourced from both the source domain X and target domain Y . The generator's encoder generates low-dimensional feature vectors. Classifier A determines whether the input image belongs to the source domain X by evaluating the feature maps generated from both the input source domain and target domain images after they have been encoded by the generator. Additionally, the class activation map (CAM) technique can calculate the weight values W for each channel in the encoded feature map using global pooling. By employing the principles of CAM, the feature map with an attention mechanism can be obtained by multiplying and summing the weights of each channel with the encoded feature map. Subsequently, this feature map with an attention mechanism is fed into the generator's decoder, where it is upsampled to restore its dimensions to match those of the input image. Classifier A within the generator serves as a binary classifier, discerning the feature maps generated from both the input source domain and target domain images.

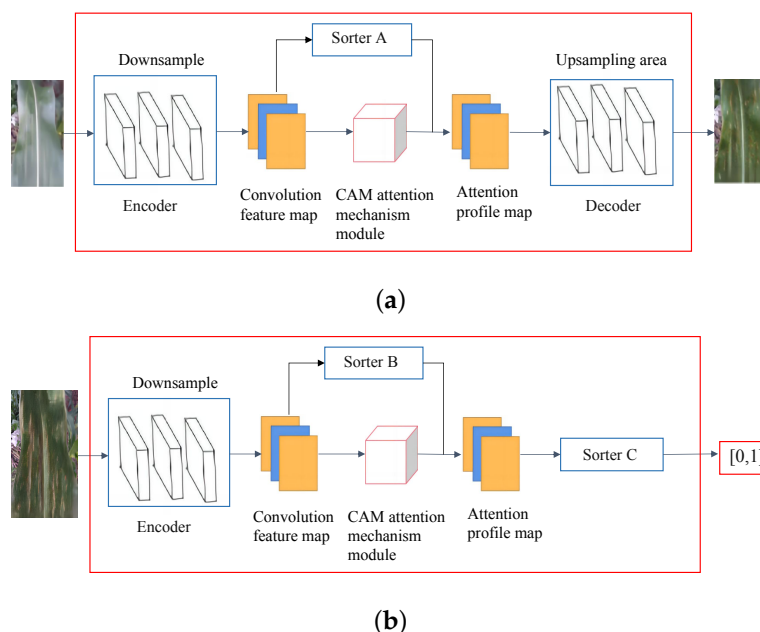


Figure 3. Improved generator and discriminator structure: (a) improved generator structure, (b) improved discriminator structure.

2.2.3. Discriminator Network Architecture

The input of the discriminator consists of both real maize disease leaf images and generated ones. Following downsampling by the encoder, the input undergoes one-dimensional convolution, combined with a bias vector, and is subsequently passed through a sigmoid function to accomplish the binary classification task utilizing probability values. A discriminator output value of 1 signifies the identification of the input maize disease image as real, while a value of 0 denotes the classification of the maize disease image as generated.

Figure 3b displays the network structure of the discriminator after integrating the CAM soft attention mechanism module into the original model.

The discriminator incorporates the soft attention mechanism module of CAM. As illustrated in Figure 3b, the discriminator receives input images from both the generated image $G(X)$ and the target domain images Y . The discriminator is equipped with two binary classifiers: classifier B and classifier C. Classifier B determines whether the input image is a generated image or an image from the target domain by processing the feature maps that are extracted from the input-generated image and the target domain image, both encoded by the discriminator D_s . The output of classifier B is the probability of the input image being a generated image. Classifier C, on the other hand, determines the same by processing the attention mechanism feature map generated via CAM. This feature map is obtained from the input-generated image and the target domain image after being processed by CAM. The output of classifier C also represents the probability of the input image being a generated image. Both classifiers B and C aim to determine whether the input image originates from a generated image or an image from the target domain.

2.2.4. Designing Feature Recombination Loss Function

The task of generating images of diseased maize leaves can be considered as a regression problem. To enhance the network's feature extraction capability, the attention mechanism module has been introduced. In order to achieve feature recombination, a regression loss function is incorporated. The specific expression of this loss function is as follows:

$$L = L_{cyc} + \mu L_{cyc}(G, F) + \lambda(L_A + L_B) + \beta L_f^k \quad (6)$$

In the equation, the loss of the improved CycleGAN model is denoted by L . The terms L_{cyc} and $(L_A + L_B)$ represent the adversarial losses, which aim to enhance the generative capability of the generator and the discriminative ability of the discriminator, respectively. $L_{cyc}(G, F)$ represents the cycle consistency loss, which is utilized to enforce the generalization ability of the image translation and improve the fidelity of the generated images. The term L_f^k corresponds to the regression loss, employed to improve the quality of the generated maize disease leaf images. The constants μ , λ , and β represent the respective weights of these terms in the overall loss. Adjusting the values of these three parameters during training can yield different outcomes.

The adversarial loss L_{gan} comprises two GAN network adversarial losses: $L_{gan}(G, D_Y, X, Y)$ and $L_{gan}(G, D_X, Y, X)$. These losses are essentially binary cross-entropy functions. The specific expressions for these functions are illustrated in Equations (7) and (8):

$$L_{gan}(G, D_Y, X, Y) = E_{y \sim P_{data}(y)}(\log D_Y(y)) + E_{x \sim P_{data}(x)}(\log (1 - D_Y(G(x)))) \quad (7)$$

$$L_{gan}(F, D_X, X, Y) = E_{x \sim P_{data}(x)}(\log F_X(x)) + E_{y \sim P_{data}(y)}(\log (1 - F_Y(G(y)))) \quad (8)$$

Equation (7) depicts the generation process from healthy maize leaf images to diseased maize leaf images. In this equation, y represents the input real maize disease leaf image, while x denotes the generated maize disease leaf image. $D_Y(y)$ signifies the discriminator's probability of classifying the input as a real maize disease leaf image, whereas $D_Y(G(x))$ indicates the discriminator's probability of classifying the input as a generated maize disease leaf image.

Equation (8) illustrates the generation process from diseased maize leaf images to healthy maize leaf images. In this equation, x refers to the input real healthy maize leaf image, while y represents the input generated healthy maize leaf image. $F_X(x)$ represents the discriminator's probability of classifying the input as a real healthy maize leaf image, while $F_Y(G(y))$ represents the discriminator's probability of classifying the input as a generated healthy maize leaf image.

Adversarial loss ($L_A + L_B$) consists of two binary cross-entropy loss functions from two classifiers, as shown in Equations (9) and (10):

$$L_A = -(E_{x \sim X}(\log A(x)) + E_{x \sim Y}(\log (1 - A(x)))) \quad (9)$$

$$L_B = E_{x \sim G(x)}(\log (B(x))) + E_{x \sim Y}(\log (1 - B(x))) \quad (10)$$

In Equation (9), classifier A conducts binary classification on real maize disease leaf images and fake maize disease leaf images, thereby serving as a binary classification task for the generator. On the other hand, in Equation (10), classifier B performs binary classification on generated maize disease leaf images and real maize disease leaf images, acting as a binary classification task for the discriminator.

Furthermore, the cycle consistency loss $L_{cyc}(G, F)$ measures the mean distance between predicted values and true values. This loss ensures the consistency of the translated images from one domain to another. The specific expression for this function is depicted in Equation (11):

$$L_{cyc}(G, F) = E_{x \sim Pdata(x)}(\|F(G(x)) - x\|_1) + E_{y \sim Pdata(y)}(\|G(F(y)) - y\|_1) \quad (11)$$

In Equation (11), the following variables are defined: x represents the input image of a healthy maize leaf, $G(x)$ represents the image of a maize leaf with disease generated by generator G , $F(G(x))$ represents the image of a healthy maize leaf generated by generator F using $G(x)$ as input. Additionally, y represents the input image of a maize leaf with disease, $F(y)$ represents the image of a healthy maize leaf generated by generator F using y as input, and $G(F(y))$ represents the image of a maize leaf with disease generated by generator G using the output of generator F as input.

The proposed feature recombination loss, denoted as L_f^k , measures the dissimilarity between real and generated maize disease leaf images in the hidden layer of the discriminator. This loss is calculated based on the minimum absolute error for feature differences. The specific expression for this function is presented in Equation (12):

$$L_f^k = \frac{1}{n} \sum_{i=1}^N \|F_k(y_i) - F_k(D(x_i))\| \quad (12)$$

In Equation (12), x_i represents the input representation for the i -th healthy maize leaf image, y_i represents the representation for the i -th real maize disease leaf image, k represents the corresponding hidden layer number, N represents the number of samples for comparison, D represents the generator for maize plants, $D(x_i)$ represents the conversion of the input healthy maize leaf image to a maize disease leaf image, F_k represents the representation of image features on the k -th hidden layer, and $F_k(y_i)$ and $F_k(D(x_i))$, respectively, represent the feature representations that exist in the k -th hidden layer of the discriminator for the real maize disease leaf image and the generated maize disease leaf image. The magnitude of L_f^k reflects the similarity between the real maize disease image and the generated maize disease image. That is, the smaller the L_f^k value, the closer the generated maize disease image is to the real maize disease image, implying higher quality of the generated maize disease image.

2.3. Type Training and Parameters

The training of the CycleGAN model is conducted using the TensorFlow deep learning framework and accelerated using a GPU. Residual blocks are employed in the generator to address the problem of degradation in deep neural networks, which simultaneously enhances the convergence speed of the network. The training process of the model consists of the following stages:

- (1) Pre-train the network parameters using a small custom dataset of maize disease.
- (2) Train the model using a substantial amount of custom maize disease dataset. The generator and discriminator's network parameters are trained in a step-by-step manner. Real-time monitoring of the training process is conducted using the TensorBoard module.
- (3) Fix the generator parameters and train the discriminator parameters. The discriminator is updated at a 3:1 ratio compared to the generator.
- (4) The training is considered complete when both discriminators cannot determine the source of the maize disease leaf image. This is reflected in an output value of 0.5, indicating the Nash equilibrium.

Different learning rates are compared and analyzed to examine their effect on the utilization degree of output errors. Therefore, different learning rates (0.01, 0.001, 0.0001) are chosen and the loss function of both the training set and test set are used to study the impact of the learning rate on the model.

The impact of different learning rates on the loss function is analyzed, as depicted in Figure 4a. It is observed that the loss function decreases relatively smoothly after 15,000 epochs with a learning rate of 0.001. The value of the loss function is lower than those corresponding to the learning rates 0.01 and 0.0001, suggesting that a learning rate of 0.001 is optimal for training the model. Figure 4b reveals that the loss function value exhibits relatively large fluctuations when the learning rate is set to 0.01 and 0.0001. Initially, the loss function value experiences instability but later stabilizes. However, when the learning rate is set to 0.001, the overall decrease in the loss function is similarly smooth. This learning rate also yields the optimal value for the loss function among the three learning rates considered. Based on an analysis of the training and testing loss values, it is concluded that the optimal initial learning rate for the model is 0.001. The comparison of learning rate experiments revealed that beyond 80,000 rounds, the rate of decrease becomes insignificant, and this decrease can be disregarded. Consequently, it is determined that the appropriate number of training iterations is 80,000 rounds.

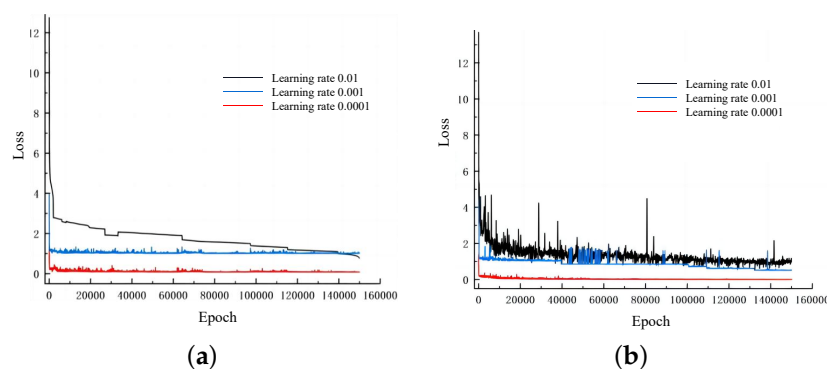


Figure 4. Different learning rates correspond to the loss values: (a) training loss value, (b) test loss value.

Both the input and output maize images in the experiment have dimensions of 256×256 pixels. For batch processing, a batch size of 50 is used, and the model is trained for 80,000 rounds. The weights are saved at intervals of 1000 rounds. The initial learning rate is set to 0.001 and gradually decreases linearly to 0 after 100,000 rounds. The model employs rectified linear unit (ReLU) as the activation function for non-linear correction and utilizes the Adam algorithm for gradient descent optimization. Table 1 presents the parameters of each module in the network.

Table 1. Improved CycleGAN parameters.

Network Module	Area	Input Dimension →Output Dimension	Number and Size of Core	Network Layer Information
Generator	Subsampled	$3*m*n$ → $64*m*n$	$7*7*64$	Convolutional layer, IN,ReLU,stride 1
		$64*m*n$ → $128*m/2*n/2$	$3*3*128$	Convolutional layer, IN,ReLU,stride 2
		$128*m/2*n/2$ → $256*m/4*n/4$	$3*3*256$	Convolutional layer, IN,ReLU,stride 2
	Residual network	$256*m/4*n/4$ → $256*m/4*n/4$	$3*3*256$	3 residual blocks,IN, ReLU,stride 1
	Attention	$256*m/4*n/4$ → $512*m/4*n/4$	$3*3*512$	Max pooling and average pooling,stride 1
		$512*m/4*n/4$ → $256*m/4*n/4$	$3*3*256$	Convolutional layer, ReLU,stride 1
	Residual network	$256*m/4*n/4$ → $256*m/4*n/4$	$3*3*256$	3 residual blocks, IN,ReLU,stride 1
	Upsampling	$256*m/4*n/4$ → $128*m/2*n/2$	$3*3*128$	Convolutional layer, IN,ReLU,stride 2
		$128*m/2*n/2$ → $64*m/2*n/2$	$3*3*64$	Convolutional layer, IN,ReLU,stride 2
		$64*m/2*n/2$ → $3*m*n$	$7*7*3$	Convolutional layer, Tanh,stride 2
Discriminator	Subsampling	$3*m*n$ → $64*m/2*n/2$	$4*4*64$	Convolutional layer, AdaIN,ReLU,stride 2
		$64*m/2*n/2$ → $128*m/4*n/4$	$4*4*128$	Convolutional layer, AdaIN,ReLU,stride 2
		$128*m/4*n/4$ → $256*m/8*n/8$	$4*4*256$	Convolutional layer, AdaIN,ReLU,stride 2
		$256*m/8*n/8$ → $512*m/16*n/16$	$4*4*512$	Convolutional layer, AdaIN,ReLU,stride 2
	Attention	$512*m/16*n/16$ → $1024*m/16*n/16$	$4*4*1024$	Max pooling and average pooling,stride 1
		$1024*m/16*n/16$ → $512*m/16*n/16$	$4*4*512$	Convolutional layer, ReLU,stride 1
	Classifier	$512*m/16*n/16$ → $1*m/16*n/16$	$4*4*1$	Convolutional layer, ReLU,stride 1

3. Experimental Process

To verify the effectiveness of the improved CycleGAN model proposed in this study, the following steps were taken: (1) First, the original dataset was constructed based on the public dataset 2018 AI challenge. Structural similarity (SSIM) [44], peak signal-to-noise ratio (PSNR), and Fréchet inception distance (FID) [45] were used as objective evaluation indicators to assess the quality of the generated images, and grayscale histograms were used as subjective evaluation indicators. (2) The stability of the improved model was compared through multiple experiments to verify its generating capability and stability. (3) An ablation experiment was conducted to evaluate the performance of the attention mechanism and feature recombination loss function proposed in this study, comparing them with the VAE and Pix2Pix image translation models. (4) The original CycleGAN model, the improved model, and the Pix2Pix model were used to expand the original dataset, and the average accuracy was compared on classification models, including VGG16, VGG19, ResNet50, DenseNet50, DenseNet121, and GoogLeNet. Confusion matrices were

established based on the accuracy of VGG16, ResNet50, DenseNet50, and GoogLeNet to provide further insights into the experimental results.

3.1. Evaluation Metrics

To verify the improved model's performance in generating images of diseased maize leaves, this study utilized four evaluation indicators to analyze the quality of the generated images. The evaluation indicators used included structural similarity (SSIM), the peak signal-to-noise ratio (PSNR), and the Fréchet inception distance (FID).

SSIM assesses the similarity between the data distribution of a healthy maize leaf image (X) and the generated diseased maize leaf image (Y).

$$\text{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_1)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (13)$$

where μ_x and μ_y are the means of images X and Y , σ_x and σ_y are the variances of images X and Y , and σ_{xy} is their covariance. Moreover, C_1 and C_2 are introduced to prevent the occurrence of division-by-zero exceptions in the formula. The range of SSIM is from -1 to 1 , with higher values indicating greater similarity in the distribution structure between the two images and higher image quality.

Peak signal-to-noise ratio (PSNR) is a metric utilized to compare the errors between corresponding pixels in the healthy maize leaf image and the generated diseased maize leaf image. The formula for PSNR is shown in Equation (14).

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{Max} \cdot W \cdot H}{\sum_i^H \sum_j^W (X(i,j) - Y(i,j))^2} \right) \quad (14)$$

Here, Max represents the maximum grayscale level of the image, which is 255. $X(i,j)$ refers to the pixel value of the original image, $Y(i,j)$ refers to the pixel value of the generated image, H represents the image height, and W represents the image width. The PSNR metric is measured in dB, where a higher value indicates better image quality.

The Fréchet inception distance (FID) is a metric that quantifies the distance between the high-dimensional data distributions of real and generated images. The FID result is a numerical value that provides an intuitive measure of the degree of similarity between the distributions. The formula for the FID is shown in Equation (15).

$$\text{FID}(x, y) = \|u_x - u_g\|_2^2 + \text{Tr}(\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{1/2}) \quad (15)$$

Here, Tr represents the trace of a matrix, which is the sum of its diagonal elements. Variable u represents the mean, and Σ_x and Σ_g represent the covariances. The real sample images are denoted as x , while the generated sample images are represented as g . The tuples (u_x, u_g) and (Σ_x, Σ_g) represent the means and covariances calculated from the real data and generated samples, respectively. A lower value of FID indicates a higher degree of similarity between the images.

3.2. Experimental Data

The experimental data utilized in this study were constructed from the publicly available 2018 AI Challenge dataset (<https://aistudio.baidu.com/datasetdetail/76075> accessed on 3 September 2023). Within this dataset, images were selected from the 2018 AI Challenger Plant Disease Degree Image Dataset, which consisted of various datasets, including healthy maize leaf images, maize gray leaf spot images, maize rust images, and maize leaf blight images. The diseased images were categorized into two levels of severity: mild and severe. In order to obtain a training set, any images that were deemed unclear or indiscernible were manually removed. The resulting training set consisted of 370 healthy maize leaf images, 191 mild and 167 severe maize gray leaf spot images, 309 mild and 227 severe maize rust images, and 113 mild and 329 severe maize leaf blight images. To overcome the insufficient number of images available for each type of maize disease, data augmentation techniques

were employed. These techniques included horizontal flipping, vertical flipping, clockwise rotation by 45 degrees, and counterclockwise rotation by 45 degrees, thereby augmenting the dataset and increasing the sample size. Subsequently, the dataset was divided into training and testing sets using a 7:3 ratio. In order to ensure consistency, the images in the public dataset, which possessed varying resolutions, were normalized and scaled to a fixed pixel size of 224×224 RGB color images, as required by the model. This normalization step was necessary to address the redundancy in the images caused by a high number of pixels. Additionally, the images were further processed to create binary files for the image dataset. The entire process of experimental data preparation was completed accordingly.

4. Experimental Results and Analysis

4.1. The Impact of the Improved Model on Model Performance

Both the generator and discriminator models employed in this study incorporated an attention mechanism structure. Additionally, a feature recombination loss function was introduced to the discriminator. The inclusion of these enhancements facilitated a deeper network architecture and improved the network's capability to extract features from deeper layers, surpassing the performance of the original CycleGAN model. Figure 5 and Figure 6 present a visual comparison of the image quality between the original CycleGAN and the improved CycleGAN for both mild and severe maize disease leaf images, respectively.

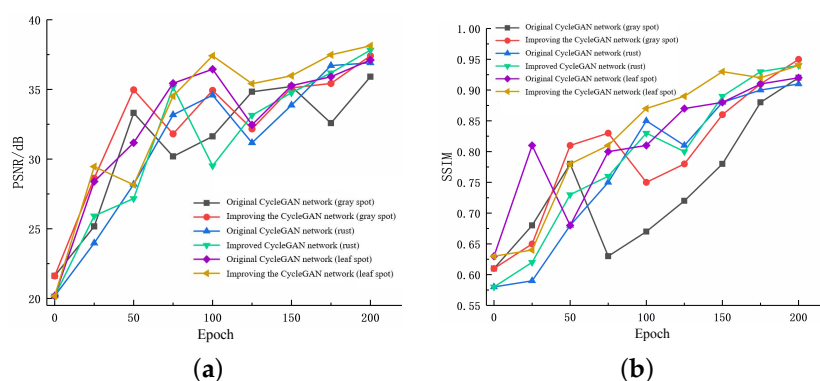


Figure 5. Comparison of the original and improved networks that generated leaf images of minor maize disease: (a) the change in PSNR value, (b) the change in PSNR value.

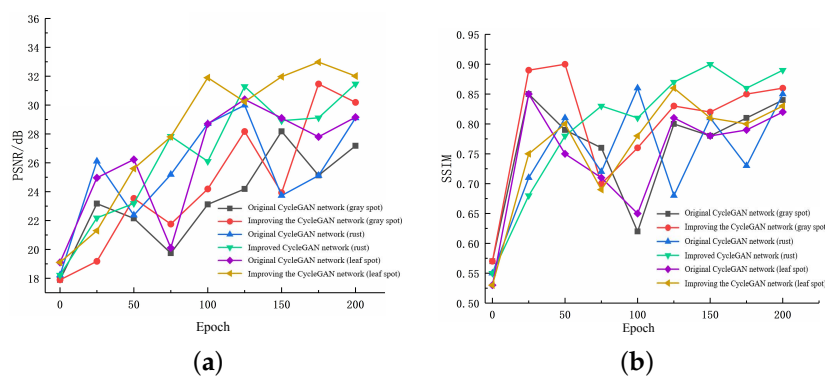


Figure 6. Comparison of the original and improved networks that generated leaf images of severe maize disease: (a) the change in PSNR value, (b) the change in PSNR value.

From Figure 5, it is evident that the two models generated three types of mild maize leaf images. For the maize gray leaf spot image, the original CycleGAN yielded PSNR and SSIM values of 35.91 dB and 0.92, respectively, while the improved CycleGAN achieved values of 37.41 dB and 0.95, resulting in a 2.50 dB increase in PSNR and a 0.03 increase in SSIM. Similarly, for maize rust leaf images, the original CycleGAN produced PSNR and SSIM values of 36.91 dB and 0.91, respectively, while the improved CycleGAN achieved

values of 37.82 dB and 0.94, representing a 1.71 dB increase in PSNR and a 0.03 increase in SSIM. Lastly, for maize leaf blight images, the original CycleGAN yielded PSNR and SSIM values of 37.11 dB and 0.92, respectively, while the improved CycleGAN achieved values of 38.13 dB and 0.94, resulting in a 1.02 dB increase in PSNR and a 0.02 increase in SSIM.

From Figure 6, it is evident that the two models generated three types of severe maize leaf images. For the maize gray leaf spot image, the original CycleGAN yielded PSNR and SSIM values of 27.18 dB and 0.84, respectively, while the improved CycleGAN achieved values of 30.19 dB and 0.86, resulting in a 3.01 dB increase in PSNR and a 0.02 increase in SSIM. Similarly, for maize rust leaf images, the original CycleGAN produced PSNR and SSIM values of 29.11 dB and 0.85, respectively, while the improved CycleGAN achieved values of 31.47 dB and 0.89, representing a 2.36 dB increase in PSNR and a 0.04 increase in SSIM. Lastly, for maize leaf blight images, the original CycleGAN yielded PSNR and SSIM values of 29.16 dB and 0.82, respectively, while the improved CycleGAN achieved values of 32.01 dB and 0.83, resulting in a 2.85 dB increase in PSNR and a 0.01 increase in SSIM.

By examining Figures 5 and 6, it is evident that the generation process of mild maize disease leaf images exhibits relatively small fluctuations in the PSNR and SSIM curves for both the original CycleGAN and the improved CycleGAN. However, when generating severe maize disease leaf images, both models experience larger fluctuations in the PSNR and SSIM curves, which gradually stabilize as the number of iterations increases. Overall, the proposed improved CycleGAN structure in this study has achieved higher PSNR and SSIM values for the three different disease severity levels of maize diseases compared to the original CycleGAN, indicating superior quality in the generated images.

4.2. Contrast Based on Objective Parameters of Generated Images

(1) Comparison of generated image FID values.

The objective parameter, FID, is used for the preliminary assessment of the generated image quality, providing an intuitive and effective evaluation. Visual observation allows for the identification of similarities and differences between generated and real images, providing preliminary judgments on the generated image performance based on parameter similarities. Subsequently, the distribution similarity between generated and real images can be assessed. Table 2 displays the FID values generated for three different levels of maize disease. Model A represents a network that solely integrates the attention mechanism, while model B represents a network that solely integrates feature recombination. Conversely, model C represents a network that combines both the attention mechanism and feature recombination.

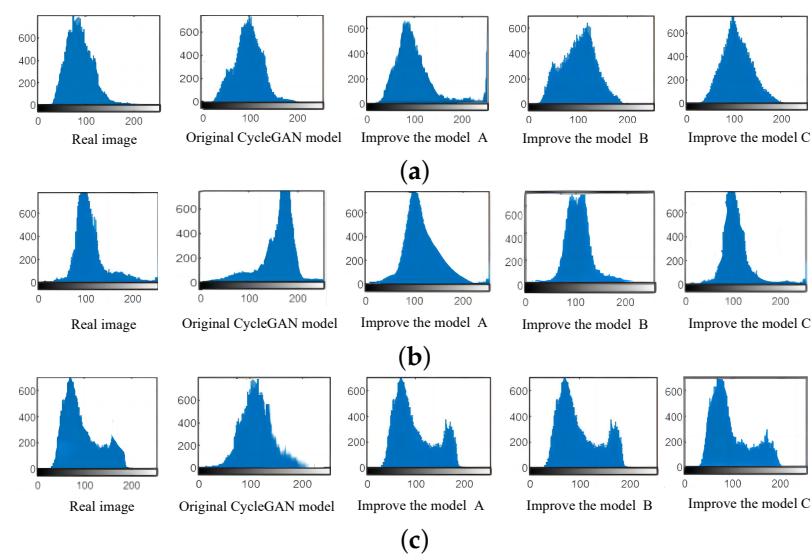
From Table 2, the results show that, in terms of the maize gray spot disease leaf image, the improved model C generates an image that is closest to the real one based on objective parameters. Its FID value is reduced by 47.42 and 47.46 compared to the image generated by the original CycleGAN. For the maize leaf rust disease leaf image, the improved model C generates an image that is closest to the maize gray spot disease leaf image based on objective parameters, with reductions in the FID value by 57.61 and 52.43. Regarding the maize common rust disease leaf image, the improved model C generates an image that is closest to the maize gray spot disease leaf image in terms of objective parameters, with reductions in the FID values by 52.96 and 51.16.

(2) Gray-level histogram feature maps comparison.

The gray-level histogram is a statistical function that represents the distribution of gray levels in an image. It shows the frequency of occurrence of different gray levels by indicating the number of pixels with a specific gray level. Figure 7 presents the comparison of gray-level histograms for the target images generated from real images, the original CycleGAN model, and improved models A, B, and C. The horizontal axis represents the pixel value, and the vertical axis represents the number of pixels. This comparison provides insight into the variations in gray-level distributions among the different image generation models.

Table 2. FID value comparisons of maize disease leaf images generated by different models.

Module	CycleGAN	The Attention-Based CycleGAN Method	The Feature Recombination-Based CycleGAN Method	The Incorporating Attention Mechanism and Feature Recombination CycleGAN Method
gray leaf spot (minor/severe)	158.26/172.17	114.26/128.61	124.71/136.19	110.84/124.71
rust disease (minor/severe)	167.34/181.59	113.71/135.86	119.93/141.03	109.73/131.16
leaf spot disease (minor/severe)	161.32/175.99	112.36/127.41	127.53/139.76	108.36/124.83

**Figure 7.** Comparison of grayscale features of diseased maize leaf images generated by different models: (a) maize gray leaf spot disease, (b) maize rust, (c) maize leaf spot.

The results indicate that, based on Figure 7, improved model C shows a higher similarity to real maize leaf images in terms of gray level, compared to the original CycleGAN model, improved model A, and improved model B for all three types of maize disease leaf images. In conclusion, by analyzing the FID values and visually comparing gray-level histograms of the generated maize disease leaf images using objective parameters, it can be concluded that the maize disease leaf images generated by improved model C exhibit higher image quality and are more similar to real maize leaf images.

4.3. Comparison of Stability of Improved Models

In this study, a recombined feature loss function was introduced to optimize the edge information of leaf contours and enhance image generation quality. To validate the performance improvement brought by the recombined feature loss function, maize gray spot disease leaf images were used as a case study, and the image generation quality was compared for images with different background complexities. Figure 8 visually demonstrates the comparison results.

The feature contours of the generated images more closely resemble the original leaf images after adding the feature recombination module, as depicted in Figure 8. This improvement can be attributed to the module's ability to extract deep features from both diseased and healthy leaf images, enabling the improved model to better learn the mapping relationship between the two types of images. Under the condition of a simple background, the generated images using improved model B achieve PSNR and SSIM values of 23.17 dB

and 0.89, respectively, while those using improved model C achieve 25.63 dB and 0.91, respectively. Under the condition of a complex background, the generated images using improved model B achieve PSNR and SSIM values of 21.89 dB and 0.82, respectively, while those using improved model C achieve 23.11 dB and 0.86, respectively. Through the comparison of different backgrounds, it can be observed that the generated images of diseased maize leaves using improved model C attain higher PSNR and SSIM values, indicating that the generated images more closely resemble the real images in terms of pixel values and exhibit structures that more closely resemble the original images, ultimately leading to higher image quality.

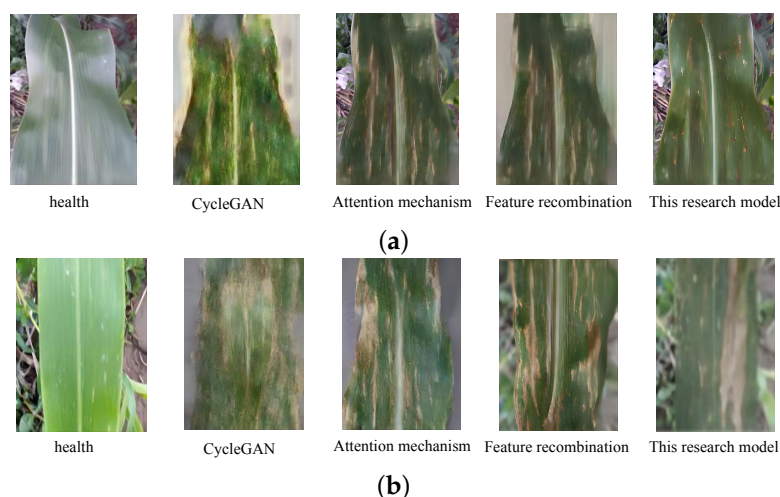


Figure 8. Improved model generation image effects in different degrees of complexity backgrounds: (a) simple background, (b) complex background.

By analyzing Figures 5–8, it is evident that improved model C exhibits satisfactory performance in generating maize disease leaf images, suggesting that the inclusion of the feature recombination loss function has demonstrated partial improvement in the training effectiveness of improved model C. Nevertheless, the feature recombination loss function exhibits instability. To ensure that the introduced feature recombination module has a better effect on the quality of generated maize disease leaf images, the analysis focuses on maize gray spot disease leaf images in six repeated experiments using identical parameters for all experimental groups.

According to Table 3, which presents the results of six repeated experiments, the mean and variance of PSNR values for leaves affected by maize gray spot disease are 23.13 dB and 0.0083, respectively. For SSIM values, their mean and variance are 0.89 and 0.0011, respectively. In the case of leaves with severe symptoms, the mean and variance of PSNR values are 20.89 dB and 0.0012, respectively. The mean and variance of SSIM values for these leaves are 0.81 and 0.0028, respectively. By comparing the mean and variance of PSNR and SSIM values, it can be observed that the variance is small and the overall data shows little fluctuation. This indicates that the introduced feature recombination loss function is not limited to achieving the best performance in a single experiment and that the overall stability of the model is not significantly affected.

Table 3. Improved Model stability evaluation.

Method	PSNR/dB		SSIM	
	Mean	Variance	Mean	Variance
minor	23.13	0.0083	0.89	0.0011
severe	20.89	0.0012	0.81	0.0028

4.4. Ablation Experiment and Comparison with Other Methods

To compare the improvement effects of the proposed method in this study, we performed controlled experiments using traditional generative models VAE and deep neural network method Pix2Pix as well as conducted ablation experiments. The experiments focused on maize gray leaf spot disease, and the quality of the generated images was compared. We compared the performance of the proposed method, original CycleGAN, attention-based CycleGAN, feature recombination-based CycleGAN, VAE, and Pix2Pix on a constructed dataset of maize gray leaf spot disease. The comparison of the six methods in generating diseased leaf images can be seen in Figure 9, and the quality comparison is presented in Table 4.

Table 4. Image quality comparison of maize disease leaves generated by different methods.

Method	PSNR/dB		SSIM	
	Minor	Severe	Minor	Severe
This research method	23.13	20.89	0.89	0.81
The original CycleGAN method	18.47	14.11	0.81	0.73
The attention-based CycleGAN method	21.83	18.64	0.86	0.78
The feature recombination-based CycleGAN method	20.02	16.91	0.83	0.75
The VAE (Variational Autoencoder) method	12.11	10.38	0.71	0.67
The Pix2Pix method	22.76	20.17	0.88	0.79

Table 4 shows that the proposed method achieves higher PSNR and SSIM values compared to the original CycleGAN, VAE, and Pix2Pix methods for generating images of both slightly and severely diseased maize gray leaf spots. In the case of slightly diseased images, the proposed method achieves a PSNR value of 23.13 dB and an SSIM value of 0.89, which are improvements of 4.66 dB and 0.08, respectively, compared to CycleGAN, 11.02 dB and 0.18 compared to VAE, and 0.37 dB and 0.01 compared to Pix2Pix. For severely diseased images, the proposed method achieves a PSNR value of 20.89 dB and an SSIM value of 0.81, which are improvements of 6.78 dB and 0.08, respectively, compared to CycleGAN, 10.51 dB and 0.14 compared to VAE, and 0.72 dB and 0.02 compared to Pix2Pix. These results demonstrate the superior performance of the proposed method in generating maize gray leaf spot disease images, which closely resemble real images.

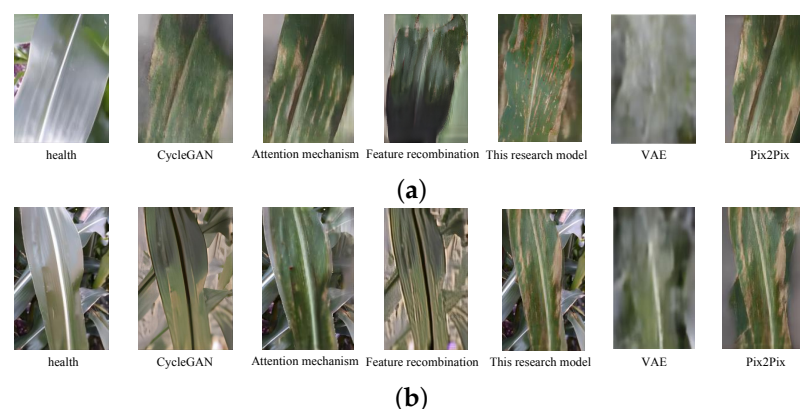


Figure 9. The comparative study of generating slightly diseased image effects using six different methods under varying degrees of complexity in different backgrounds: (a) simple background, (b) complex background.

In terms of the visual results of the generated slightly diseased maize gray leaf spot images by the six compared methods, the proposed method in this article performs the best. When generating diseased leaf images under different degrees of background complexities, the proposed method avoids generating interference images with disease features on the interfering leaves in the background and successfully repairs the edge images of maize

leaves, resulting in more realistic generated diseased images. Among the six methods compared, Pix2Pix is the closest to the proposed method in performance, but its handling of background information shows some residual information. The VAE method produces the lowest image quality, with distorted slightly diseased maize gray leaf spot images and poor clarity. The attention-based and feature recombination-based CycleGAN methods both make certain improvements in image clarity and interference information in image backgrounds compared to the original CycleGAN method, but may still struggle to fully generate disease feature information when dealing with complex backgrounds.

Due to the visually intuitive manifestation of disease severity in severe maize gray leaf spot images, it can be observed from Figure 10 that the method proposed in this paper shows better results in generating images of leaves severely affected by corn gray leaf spot disease. The other methods show residual background interference and partial distortion in the generated images.

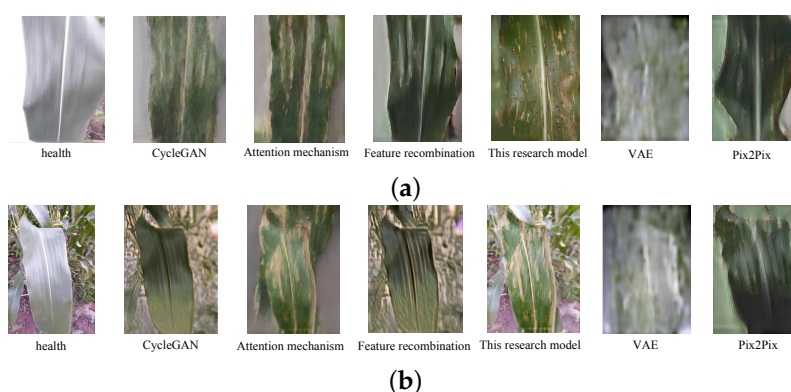


Figure 10. Results of six methods with different background complexities: (a) simple background, (b) complex background.

Experimental analysis on datasets containing both mild and severe maize gray leaf spots demonstrates the commendable performance of the proposed method in generating accurate leaf images of the disease. The generated images exhibit a higher quality, closely resembling real leaf images. Significantly, the proposed method surpasses the original CycleGAN method, as well as the attention-based and feature recombination-based CycleGAN methods. Furthermore, compared to traditional generation models like VAE and Pix2Pix, the proposed method not only improves the overall clarity of images but also effectively captures specific maize disease feature information in targeted areas. Additionally, the generated image's edge information shows a greater similarity to the original image, thus enhancing its quality.

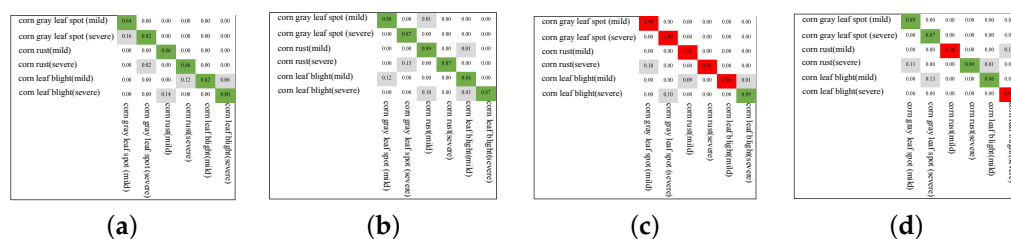
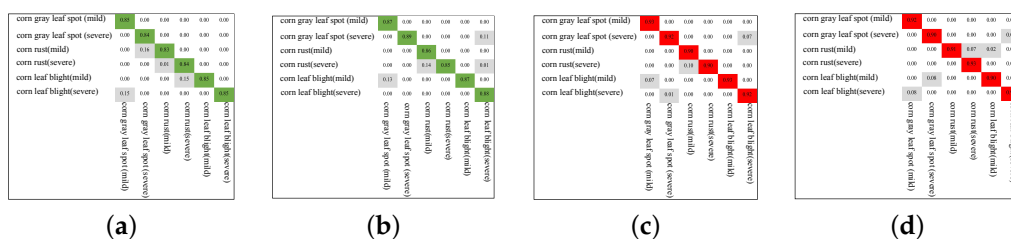
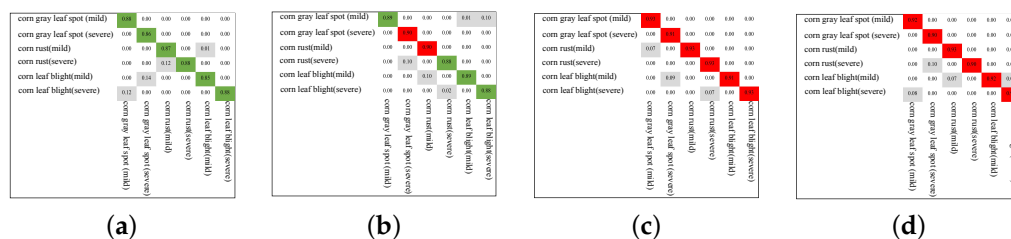
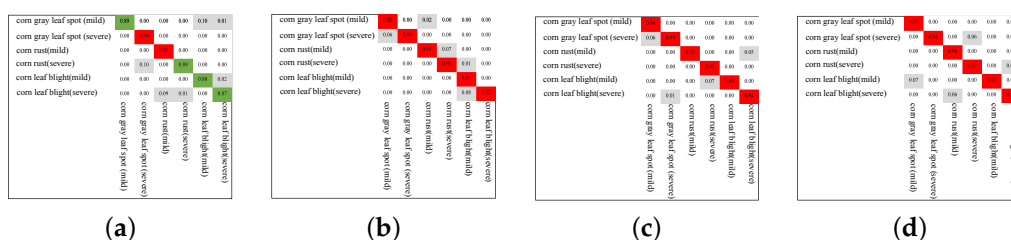
4.5. Comparison Based on Classification Model Accuracy

Based on the original base data, we employed three methods for dataset augmentation: the original CycleGAN method, the proposed method in this paper, and the Pix2Pix method. The unbalanced maize disease dataset was expanded to 500 images, and the training set and test set were divided in a 4:1 ratio, resulting in a total of 300 iterations. For batch processing, a batch size of 50 is used. The initial learning rate is set to 0.001. We calculated the average accuracy of six recognition networks (VGG16, VGG19, ResNet50, DenseNet50, DenseNet121, and GoogLeNet) using the augmented dataset, as presented in Table 5. In Table 5, A denotes the original dataset, B represents the CycleGAN model augmented dataset, C corresponds to the improved model in this paper augmented dataset, and D signifies the Pix2Pix model augmented dataset.

Table 5. Dataset average accuracy comparison in different classification models.

Model	VGG16	VGG19	DenseNet50	DenseNet121	ResNet50	GoogLeNet
A	85.19%	86.16%	84.12%	87.71%	87.27%	89.03%
B	88.13%	89.57%	87.63%	89.12%	89.12%	91.39%
C	91.11%	91.79%	92.33%	93.17%	92.93%	93.64%
D	89.94%	90.73%	90.91%	92.14%	91.68%	92.56%

Based on Table 5, the proposed improved model achieves the highest accuracy among the six classification models when generating the augmented dataset. Notably, the average accuracy using the GoogLeNet model reaches up to 93.64%, surpassing the accuracy of the original dataset, original model, and datasets expanded by deep learning generation models. For the classification results of VGG16, ResNet50, DenseNet50, and GoogLeNet, confusion matrices are presented in Figures 11–14. In these figures, (a) represents the original dataset, (b) represents the CycleGAN model augmented dataset, (c) corresponds to the augmented dataset using the improved model from this paper, and (d) signifies the Pix2Pix model augmented dataset.

**Figure 11.** Dataset in VGG16 classification results.**Figure 12.** Dataset in ResNet50 classification results.**Figure 13.** Dataset in DenseNet50 classification results.**Figure 14.** Dataset in GoogLeNet classification results.

5. Discussion

This paper introduces an enhanced CycleGAN model for transforming maize disease leaf images, thereby achieving dataset augmentation:

- (1) In the experiments, to test the influence of the improved methods on the model performance, both the original CycleGAN model and the improved CycleGAN model were used to perform disease transfer on healthy maize leaves. Overall, compared to the original CycleGAN model, the improved CycleGAN model proposed in this study achieved improvements in the PSNR and SSIM values of the three different disease severity diseases, indicating better image quality of the generated images.
- (2) We tested the impacts of different mechanisms on the improved CycleGAN model, the original CycleGAN model, the CycleGAN model with only the attention mechanism, the CycleGAN model with only the feature recombination, and the CycleGAN model incorporating both the attention mechanism and feature recombination; all models were used to perform disease transfer on healthy maize leaves. Objective parameter analysis was conducted by calculating the FID values between the generated maize disease leaf images and the original healthy maize leaf images, and visual comparisons were made using grayscale histograms. It was found that the CycleGAN model with both the attention mechanism and feature recombination generated maize disease leaf images that were closer to real maize leaf images, and the image quality was relatively better.
- (3) We validated the impact of the feature recombination loss function on the model's stability, the generated image quality was compared for maize gray leaf spot images under different background complexity levels. The experiment showed that the generated images from the model with the introduced feature recombination loss function had higher image quality, not limited to the best performance in a single experiment.
- (4) Comparative experiments were conducted through ablation experiments with the traditional generation models VAE and Pix2Pix. Maize gray leaf spot disease was taken as the experimental object, and the generated image quality was compared under different background complexity levels. The results showed that compared to the original CycleGAN method, the attention-based CycleGAN method, and the feature recombination-based CycleGAN method, the proposed method in this paper had obvious improvements. Compared with the traditional generation models VAE and Pix2Pix, this proposed method not only enhanced the overall clarity of the images, but also realized the generation of maize disease feature information in specific regions, and the edge information of the generated images was also closer to the original images.
- (5) In the experiment to validate the effectiveness of the expanded dataset, we utilized the original CycleGAN method, the improved CycleGAN, and the Pix2Pix method for dataset expansion. The experiment showed that on the expanded maize disease dataset using the improved CycleGAN, the average accuracy of the classification tasks performed by six recognition networks (VGG16, VGG19, ResNet50, DenseNet50, DenseNet121, and GoogLeNet) was the highest. The average accuracy in the GoogLeNet model reached 93.64%, and the dataset expanded by the generation model achieved an average accuracy exceeding 91% in different classification models.

Previously, Chen et al. [46] proposed an improved CycleGAN for generating synthetic samples to enhance data distribution and address issues with small-sized datasets and class imbalance. Xiao et al. [47] introduced texture reconstruction loss CycleGAN (TRL-GAN) to enhance the accuracy and generalization of the citrus greening disease identification algorithm. Liu et al. [48] designed an improved YOLOX-based tomato leaf disease identification method. They employed CycleGAN to enhance tomato disease leaf samples in the PlantVillage dataset, addressing the issue of imbalanced sample numbers. Furthermore, this study primarily expanded experiments on three types of diseases in maize leaves. Subsequent work can involve augmenting a wider variety of maize leaf disease types to

enhance the recognition accuracy of maize leaf diseases under a greater range of conditions. This method can also be extended to augment other maize disease datasets, thus enriching experimental data across various domains. In processing the experimental data used in this study, two methods, namely flipping and rotating, were employed to augment the original dataset in order to meet the data quantity requirements for the classification accuracy of the neural network model. However, beyond spatial augmentation, additional enhancement functions, such as brightness and contrast adjustment, noise injection, and color jitter were applied. These enhancement methods effectively aid in capturing underlying patterns. In future work, we will incorporate these color-related enhancement methods into the experimental section and then compare the classification performance of datasets with and without the addition of synthetic images. This will enrich the experimental data and further demonstrate the effectiveness of this approach. Additionally, due to the limited availability and poor quality of agricultural disease data, achieving sustainable development in agricultural disease data is challenging. In the future, research efforts can focus on adjusting the training epochs, exploring more suitable initial weights, and refining network architectures to address these challenges.

6. Conclusions

In this study, we improved CycleGAN by incorporating an attention mechanism module and a feature recombination loss function, thereby constructing a model for transforming maize disease leaf images. We examined the impact of different mechanisms on the enhanced CycleGAN model and evaluated the quality of generated maize gray leaf spot disease images under varying levels of background complexity. Comparative experiments were conducted through ablation experiments with traditional models. Finally, we compared the accuracy of different methods using a classification model to validate the effectiveness of the model.

The experimental results indicate that, in assessing the impact on model performance, the innovation of this study lies in comparing the generation of slightly and severely diseased maize leaf images. Taking maize gray leaf spot disease leaf images as an example, comparing the quality of generated slightly diseased maize leaf images, the improved CycleGAN showed an improvement of 2.50 dB in PSNR and 0.03 in SSIM relative to the original CycleGAN. Analyzing the quality of generated severely diseased maize leaf images, the improved CycleGAN exhibited an improvement of 3.01 dB in PSNR and 0.02 in SSIM compared to the original CycleGAN. When evaluating the FID values of generated images, for both slightly and severely diseased maize leaves, the images generated by the improved model were objectively closer to real maize gray leaf spot disease leaf images in terms of FID values. Furthermore, compared to images generated by the original CycleGAN, the FID values decreased by 47.42 and 47.46, respectively. Comparing the quality of generated images under different background complexities, the proposed improved model outperformed other methods in terms of PSNR and SSIM for both simple and complex backgrounds. Finally, using the improved CycleGAN, the maize disease dataset was expanded. The classification results of six recognition networks, including VGG16, VGG19, ResNet50, DenseNet50, DenseNet121, and GoogLeNet, confirmed the effectiveness of the generated data. After expanding the samples, the accuracy of maize disease recognition significantly improved by 3.5% to 4% compared to the original dataset and by 1.5% to 2% compared to the CycleGAN model. In summary, the model constructed in this study effectively expanded the dataset, increasing the available training samples and achieving sustainable development in agricultural disease data. The added attention mechanism enabled the model to accurately extract disease features from maize leaf disease images, enhancing the realism of generated disease images. Furthermore, by utilizing the feature reconstruction loss function, the model was able to distinguish between leaves and backgrounds, enhancing contour information in the generated images and enriching sample diversity. The results of this study can be applied to improve the early identification

and management of maize leaf diseases, ultimately contributing to increased agricultural production efficiency.

Author Contributions: Conceptualization, H.G. and M.L.; methodology, H.G. and M.L.; Software, H.G.; validation, H.G. and M.L.; formal analysis, H.G.; investigation, H.G. and R.H.; resources, H.G.; data curation, M.L. and H.L.; writing—original draft preparation, H.G., M.L. and R.H.; writing—review and editing, H.G., M.L., R.H., H.L., X.Z., C.Z., X.C. and L.G.; visualization, H.G.; supervision, H.G., X.C. and L.G.; project administration, H.G. and X.C.; funding acquisition, H.G. and X.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Jilin Scientific and Technological Development Program (20230508033RC, 20210203013SF, YDZJ202303CGZH023, 20220203133SF), the National Natural Science Foundation of China (51575367, 50775151, 42077443, U21A2040), the National Key Research and Development Program of China (2016YFD0702102).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

- Wenxia, B.; Xuefeng, H.; Gensheng, H.; Dong, L. Identification of maize leaf diseases using improved convolutional neural network. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 160–167.
- Chunshan, W.; Chunjiang, Z.; Wu, H.; Ji, Z.; Jiuxi, L.; Huaji, Z. Recognizing crop diseases using bimodal joint representation learning. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 180–188.
- Dang, M.; Meng, Q.; Gu, F.; Gu, B.; Hu, Y. Rapid recognition of potato late blight based on machine vision. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 193–200.
- Kai, S.; Zhikun, L.; Hang, S.; Chunhong, G. A research of maize disease image recognition of corn based on BP networks. In Proceedings of the 2011 third International Conference On Measuring Technology and Mechatronics Automation, IEEE, Shanghai, China, 6–7 January 2011; Volume 1, pp. 246–249.
- Kulkarni, A.H.; Patil, A. Applying image processing technique to detect plant diseases. *Int. J. Mod. Eng. Res.* **2012**, *2*, 3661–3664.
- MENG, L.; ZHONG, J.p.; LI, N. Generating Algorithm of Medical Image Simulation Data Sets Based on GAN. *J. Northeast. Univ. (Nat. Sci.)* **2020**, *41*, 332.
- Gong, M.; Chen, S.; Chen, Q.; Zeng, Y.; Zhang, Y. Generative adversarial networks in medical image processing. *Curr. Pharm. Des.* **2021**, *27*, 1856–1868. [[CrossRef](#)] [[PubMed](#)]
- Wang, C.; Wu, X.; Zhang, Y.; Wang, W. Recognition and segmentation of maize seedlings in field based on dual attention semantic segmentation network. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 211–221.
- Zhu, L.; Lu, C.; Dong, Z.Y.; Hong, C. Imbalance learning machine-based power system short-term voltage stability assessment. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2533–2543. [[CrossRef](#)]
- Batista, G.E.; Prati, R.C.; Monard, M.C. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor. Newsl.* **2004**, *6*, 20–29. [[CrossRef](#)]
- Elreedy, D.; Atiya, A.F. A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. *Inf. Sci.* **2019**, *505*, 32–64. [[CrossRef](#)]
- Chou, H.P.; Chang, S.C.; Pan, J.Y.; Wei, W.; Juan, D.C. Remix: Rebalanced mixup. In Proceedings of the Computer Vision—ECCV 2020 Workshops, Glasgow, UK, 23–28 August 2020; Proceedings, Part VI 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 95–110.
- Pei, A.; Chen, G.; Li, H.; Wang, B. Method for cloud removal of optical remote sensing images using improved CGAN network. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 194–202.
- Jin, L.; Tan, F.; Jiang, S. Generative adversarial network technologies and applications in computer vision. *Comput. Intell. Neurosci.* **2020**, *2020*, 1459107. [[CrossRef](#)] [[PubMed](#)]
- Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [[CrossRef](#)]
- Ming-Fei, H.; Xin, Z.; Jian-Wei, L. Survey on deep generative model. *Acta Autom. Sin.* **2020**, *41*, 1–34.
- Ximing, L.; Jiarun, W.; Shaoqian, W. Gans based privacy amplification against bounded adversaries. *J. Front. Comput. Sci. Technol.* **2021**, *15*, 1220.

19. Xu, Z.; Wu, S.; Jiao, Q.; Wong, H.S. TSEV-GAN: Generative Adversarial Networks with Target-aware Style Encoding and Verification for facial makeup transfer. *Knowl.-Based Syst.* **2022**, *257*, 109958. [\[CrossRef\]](#)
20. Frid-Adar, M.; Diamant, I.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **2018**, *321*, 321–331. [\[CrossRef\]](#)
21. Liu, B.; Tan, C.; Li, S.; He, J.; Wang, H. A data augmentation method based on generative adversarial networks for grape leaf disease identification. *IEEE Access* **2020**, *8*, 102188–102198. [\[CrossRef\]](#)
22. Jing, Y.; Yang, Y.; Feng, Z.; Ye, J.; Yu, Y.; Song, M. Neural style transfer: A review. *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 3365–3385. [\[CrossRef\]](#)
23. Andreini, P.; Bonechi, S.; Bianchini, M.; Mecocci, A.; Scarselli, F. Image generation by GAN and style transfer for agar plate image segmentation. *Comput. Methods Programs Biomed.* **2020**, *184*, 105268. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
25. Bulat, A.; Yang, J.; Tzimiropoulos, G. To learn image super-resolution, use a gan to learn how to do image degradation first. In Proceedings of the European Conference on Computer Vision, (ECCV), Munich, Germany, 8–14 September 2018, pp. 185–200.
26. Zen, G.; Sangineto, E.; Ricci, E.; Sebe, N. Unsupervised domain adaptation for personalized facial emotion recognition. In Proceedings of the 16th International Conference on Multimodal Interaction, Istanbul, Turkey, 12–16 November 2014; pp. 128–135.
27. Fanny; Cenggoro, T.W. Deep learning for imbalance data classification using class expert generative adversarial network. *Procedia Comput. Sci.* **2018**, *135*, 60–67. [\[CrossRef\]](#)
28. Liu, W.; Luo, Z.; Li, S. Improving deep ensemble vehicle classification by using selected adversarial samples. *Knowl.-Based Syst.* **2018**, *160*, 167–175. [\[CrossRef\]](#)
29. Brock, A.; Donahue, J.; Simonyan, K. Large scale GAN training for high fidelity.natural image synthesis. *arXiv* **2018**, arXiv:1809.11096
30. Chen, F.; Zhu, F.; Wu, Q.; Hao, Y.; Wang, E.; Cui, Y. A survey about image generation with generative adversarial nets. *Chin. J. Comput.* **2021**, *44*, 347–369.
31. Kaneko, T. Generative adversarial networks: Foundations and applications. *Acoust. Sci. Technol.* **2018**, *39*, 189–197. [\[CrossRef\]](#)
32. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
33. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**. arXiv:1411.1784.
34. 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: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
35. Lin, H.; Ren, S.; Yang, Y.; Zhang, Y. Unsupervised image-to-image translation with self-attention and relativistic discriminator adversarial networks. *Zidonghua Xuebao/Acta Autom. Sin.* **2021**, *47*, 2226–2237.
36. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
37. Zhang, J.; Sun, X.; Chen, Y.; Duan, Y.; Wang, Y. Single-Image Defogging Algorithm Based on Improved Cycle-Consistent Adversarial Network. *Electronics* **2023**, *12*, 2186. [\[CrossRef\]](#)
38. Lu, L.; Liu, W.; Yang, W.; Zhao, M.; Jiang, T. Lightweight corn seed disease identification method based on improved shufflenetv2. *Agriculture* **2022**, *12*, 1929. [\[CrossRef\]](#)
39. Hu, L.; Zhang, Y. Facial image translation in short-wavelength infrared and visible light based on generative adversarial network. *Acta Opt. Sin.* **2020**, *40*, 0510001.
40. Li, Q.; Bai, W.; Niu, J. Underwater image color correction and enhancement based on improved cycle-consistent generative adversarial networks. *Acta Autom. Sin.* **2020**, *46*, 1–11.
41. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
42. Kim, T.; Cha, M.; Kim, H.; Lee, J.K.; Kim, J. Learning to discover cross-domain relations with generative adversarial networks. In Proceedings of the International Conference on Machine Learning, PMLR, Fort Lauderdale, FL, USA, 20–22 April 2017; pp. 1857–1865.
43. Yi, Z.; Zhang, H.; Tan, P.; Gong, M. Dualgan: Unsupervised dual learning for image-to-image translation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2849–2857.
44. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [\[CrossRef\]](#)
45. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
46. Chen, Y.; Pan, J.; Wu, Q. Apple leaf disease identification via improved CycleGAN and convolutional neural network. *Soft Comput.* **2023**, *27*, 9773–9786. [\[CrossRef\]](#)

47. Xiao, D.; Zeng, R.; Liu, Y.; Huang, Y.; Liu, J.; Feng, J.; Zhang, X. Citrus greening disease recognition algorithm based on classification network using TRL-GAN. *Comput. Electron. Agric.* **2022**, *200*, 107206. [[CrossRef](#)]
48. Liu, W.; Zhai, Y.; Xia, Y. Tomato Leaf Disease Identification Method Based on Improved YOLOX. *Agronomy* **2023**, *13*, 1455. [[CrossRef](#)]

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