

Research Article

Design of Painting Art Style Rendering System Based on Convolutional Neural Network

Xingyu Xie and Bin Lv 

School of Art, East China University of Technology, Nanchang, Jiangxi 330000, China

Correspondence should be addressed to Bin Lv; 201260007@ecut.edu.cn

Received 4 August 2021; Revised 22 August 2021; Accepted 11 September 2021; Published 4 October 2021

Academic Editor: Bai Yuan Ding

Copyright © 2021 Xingyu Xie and Bin Lv. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Convolutional Neural Network- (CNN-) based GAN models mainly suffer from problems such as data set limitation and rendering efficiency in the segmentation and rendering of painting art. In order to solve these problems, this paper uses the improved cycle generative adversarial network (CycleGAN) to render the current image style. This method replaces the deep residual network (ResNet) of the original network generator with a dense connected convolutional network (DenseNet) and uses the perceptual loss function for adversarial training. The painting art style rendering system built in this paper is based on perceptual adversarial network (PAN) for the improved CycleGAN that suppresses the limitation of the network model on paired samples. The proposed method also improves the quality of the image generated by the artistic style of painting and further improves the stability and speeds up the network convergence speed. Experiments were conducted on the painting art style rendering system based on the proposed model. Experimental results have shown that the image style rendering method based on the perceptual adversarial error to improve the CycleGAN + PAN model can achieve better results. The PSNR value of the generated image is increased by 6.27% on average, and the SSIM values are all increased by about 10%. Therefore, the improved CycleGAN + PAN image painting art style rendering method produces better painting art style images, which has strong application value.

1. Introduction

In recent years, deep learning has been widely used in many fields such as medical imaging [1], remote sensing [2], and three-dimensional modeling [3] and has played an important role in promoting the application of artificial intelligence in multiple industries. In order to discover useful macro information in the data, the purpose of deep learning is to combine low-level features to form more abstract features with strong representation ability. There are many types of image styles, such as oil painting, ink painting, sketch painting, etc., and applying these styles to a given image increases the diversified presentation of the image. The use of deep learning methods for style rendering is one of the hotspots in the field of image research. Image style rendering is an image conversion method based on deep learning, which can be widely used in image processing,

computer picture synthesis, and computer vision. The original image style rendering is based on the optimization method proposed in [4], which uses the backpropagation of the convolutional neural network (CNN) and uses pixel-by-pixel comparison to obtain the optimal image conversion model.

Style rendering is a processing method for rendering the semantic content of an image in different styles. The style is extracted from the specified style image template x , and the extracted style features are mapped to y through mapping T without destroying the content of the image y . y performs style conversion; that is, it realizes the image style rendering of $T(x) \approx y$. The study in [5] combined deep learning and texture generation methods, using CNN to represent style images as content feature and style feature. The high-level convolutional layer of CNN model extracts global features of image content, and the low-level convolutional layer

describes the image style. This method can generate a more objective artistic style effect. However, this method has slow convergence speed, long rendering time, and sometimes poor style expression. Image stylization methods based on model iteration [6] include two types of methods based on generative models and image reconstruction decoders. The study in [7] used the perceptual loss function to train the feedforward network and used the VGGNet trained on ImageNet to simplify the loss function calculation process and directly generate stylized images. The efficiency was improved by three orders of magnitude. At the same time, the pixel-by-pixel difference loss function is improved to a perceptual loss function, and super-resolution reconstruction is efficient, but this method requires artificially constructing a complex loss function. The study in [8] proposed an image stylization method based on a conditional adversarial network, which does not require artificial construction of loss functions and mapping functions, which simplifies the image stylization process.

The network structure of the generative adversarial network (GAN) does not need to construct a complex loss function and achieves the global optimum through the min-max game process of the generator G and the discriminator D [9]. The conditional generative adversarial network (CGAN) uses different learning models. The GAN model is a mapping $G: z \rightarrow y$ from the random noise vector z to the output image y , while the conditional GAN model is to learn from the random noise vector z and observe the image. The generative model of the mapping is from x to the output image y , namely, $G: (x, z) \rightarrow y$. The GAN style rendering effect is better than the CGAN model, but there is a problem that the coloring effect is not ideal.

Academia researches have proposed many improved image style conversion methods based on the GAN model. The study in [10] proposed to add feature space loss (L_{fea}) and image space loss (L_2) when training GAN. The study in [11] took the feature map difference of the middle layer of the network as the perceptual loss function and used the style transfer and superdivision images obtained by GAN to achieve real-time stylization and four times of super-resolution. The study in [12] obtained the super-resolution GAN model from low-resolution images to super-resolution images based on GAN. The study in [13] proposed a perceptual adversarial network (PAN) model that combines perceptual loss and GAN model and realized a variety of image style conversion applications. In the research of image style rendering, the method in [14] has been proved to be able to obtain good results, but its perceptual loss network is a pretrained VGG-16 network. The loss in the network is not easy to optimize, and the network is mainly used for classification. Although the ability to recognize the subject of the image is stronger, the ability to retain the background is weak. In view of this, this paper proposes a GAN model training method based on the PAN model. This method combines adversarial loss, content loss, and style loss into a new perceptual loss function, which can alternately update the loss network and the image style conversion network, thereby replacing the fixed loss network [15].

In view of the abovementioned image stylization problems, this paper proposes to use a new type of cycle

GAN model (CycleGAN) to achieve image style rendering and adopts a PAN model-based confrontation training method for CycleGAN model. The proposed method combines perceptual loss, content loss, and style loss into a new perceptual loss function, and the loss network and the image style conversion network can be alternately updated, thereby replacing the fixed loss network [16] and at the same time improving the original generator network structure. The experimental results prove that the proposed method can enhance the background definition of the image, make it closer to the original image in content and style, and at the same time increase the convergence speed, and the generated style rendering effect is more realistic.

The rest of the paper is organized as follows: Section 2 gives the related works about painting style rendering system based on CNN models, including conventional generative adversarial network, the perceptual adversarial network, and the image style rendering process. Section 3 illustrates the proposed improved CycleGAN + PAN model, image conversion network, and discriminator network. Section 4 gives the novel proposed combination of loss functions, such as content loss function, style loss function, and perceptual adversarial loss function. Moreover, the painting art segmentation rendering system is also given. Section 5 gives the painting art style rendering experiments and results on four different painting art forms. Section 6 concludes the model and results of the paper.

2. Related Works

2.1. Generative Adversarial Network. In the process of generating painting style images in the GAN model, the generator G constantly learns and improves the ability to produce image data, and the discriminator D gradually improves the ability to discriminate data, and G and D reach a balance in the process of confronting the game. The adversarial relationship between G and D_Y in GAN is shown in the following equation:

$$\min_G \max_{D_Y} V(D_Y, G) = E_{y \in P_{\text{data}}(y)} [\log D_Y(y)] + E_{z \in P_z} [\log(1 - D(G(z)))], \quad (1)$$

where y represents the image data, z is the noise input to the G network, and $G(z)$ is the generated image. $P_{\text{data}}(y)$ is the distribution of the real image data, $P_z(y)$ represents the noise distribution of the input network G , and \Leftrightarrow represents the data obeys the distribution relationship. E is the mathematical expectation function. The deep convolutional GAN model (DCGAN) [17] is a combination of CNN and GAN, which generates images randomly, and cannot meet the requirements of style rendering between images. Therefore, this paper selects the CycleGAN model that is based on the CNN model to construct the painting art style rendering system.

2.2. Perceptual Adversarial Network. Inspired by the GAN model, on the basis of the existing perceptual loss [18] research, reference [19] proposed the perceptual adversarial

network (PAN) model. PAN model combines perceptual loss and generative adversarial loss and conducts adversarial training between the image style conversion network and the discriminant network. Such model can continuously and automatically find the difference between the output and the real image that has not been reduced. Therefore, when the difference measured in the high-dimensional space is small, the hidden layer of the discriminant network model will be updated to a higher level, and the difference between the new output image and the real image will continue to be searched for until converging to the optimal solution.

The innovation of PAN is that there is no longer a need for a complex loss function constructed based on human experience like traditional image models. This method automatically learns the mapping from input to output pictures through the adversarial network and applies it to the image conversion problem to achieve a generalization model. The PAN model is based on GAN model, which is combined with perceptual loss for adversarial training, and enhances the naturalness and realism of the image. The PAN model can realize a variety of image conversion tasks, such as image super-resolution, denoising, semantic segmentation, automatic completion, etc. Therefore, in this paper, we used the PAN model to improve the performance and efficiency of the CycleGAN model for rendering painting art style images.

2.3. Image Style Rendering Process. The image style rendering process is divided into two stages: training phase and performing phase [20]. In the training phase, the system selects a style map Y_s and trains a transformation network model based on each style map. The samples of the content map X are continuously obtained from the training set through iterations. In each iteration, the conversion style network converts the content image X to Y and randomly transports X and Y to the discriminant network D . The network D judges between X and Y (content) and Y_s and Y (style) through the adversarial loss function, and the difference between the two is fed back to the network T . T adjusts the weights and parameters and enters the next iteration. At the same time, the network D is continuously optimized to find more differences. The final purpose is to generate an image conversion network model with Y style. In the execution stage, the system inputs any content map into the well-trained Y style conversion model and converts the content map into a Y style effect image in real time, and the original content and structure remain unchanged. Therefore, this paper improves the image painting art style rendering network, as shown in Figure 1.

3. Improved Image Style Rendering Network Structure

Although the VGG-16 loss network can be well trained with the ImageNet data set, the loss in the VGG-16 network is not easy to optimize. If a supervision item can be added to the hidden layer of D to measure the effect of generation, the loss

can be changed at any time when the network T is updated. Based on this setting, it is easier to get a better T .

3.1. Cycle Generative Adversarial Network. The CycleGAN model contains two generators and discriminators to realize the mutual mapping between images x and y [21]. Suppose the mapping of $X \rightarrow Y$ is the generator F , and the image sample $x \in X$ generates an image similar to the sample $y \in Y$ through $F(x)$. CycleGAN uses the discriminator D_Y to determine whether the generated image is a real image or not. The loss function L between the generator F and the discriminator D_Y is defined as

$$L_{\text{GAN}}(F_{X \rightarrow Y}, D_Y) = E_{y \in P_{\text{data}}(y)} [\log D_Y(y)] + E_{x \in P_{\text{data}}(x)} [\log(1 - D_Y(F_{X \rightarrow Y}(x)))], \quad (2)$$

where \Leftrightarrow , P_{data} , and E are defined with equation (1). The single adversarial loss function L cannot guarantee that the learning function maps the input x to the expected output y ; that is, it is not possible to train GAN only with equation (2). CycleGAN introduces the mapping from Y to X as $G_{Y \rightarrow X}$, and the discriminator D_X judges whether the image similar to x generated by $y \in Y$ through $G(y)$ is true or not. We can deduce the mapping loss function L between $G_{Y \rightarrow X}$ and the discriminator D_X similar to equation (2), as shown in the following equation:

$$L_{\text{GAN}}(F_{X \rightarrow Y}, D_X) = E_{x \in P_{\text{data}}(x)} [\log D_X(x)] + E_{y \in P_{\text{data}}(y)} [\log(1 - D_X(G_{Y \rightarrow X}(y)))]. \quad (3)$$

CycleGAN introduces the loss function of cycle consistency loss (CCL) and learns the two mappings $F_{X \rightarrow Y}$ and $G_{Y \rightarrow X}$ at the same time. After x is converted to y , from y to \hat{x} , the loss between x and \hat{x} is calculated to avoid, so that it is possible that all images in X are mapped to the same image in Y . Figure 2 shows the training process of the cycle consistency loss function.

Figure 2 shows the forward consistency and backward consistency of CycleGAN model. The forward consistency and backward consistency constitute the overall consistency of CycleGAN model. The overall consistency is the key of the CycleGAN model. In this paper, we use $F(G(y)) \approx y$ and $G(F(x)) \approx x$ to define the cycle consistency loss L_{CCL} as

$$L_{\text{CCL}}(F_{X \rightarrow Y}, G_{Y \rightarrow X}) = E_{x \in P_{\text{data}}(x)} (\|G_{Y \rightarrow X}(F_{X \rightarrow Y}(x)) - x\|_1) + E_{y \in P_{\text{data}}(y)} (\|F_{X \rightarrow Y}(G_{Y \rightarrow X}(y)) - y\|_1). \quad (4)$$

From equations (2)–(4), the objective function L' of CycleGAN shown in equation (5) is obtained:

$$L'(F_{X \rightarrow Y}, G_{Y \rightarrow X}) = L_{\text{GAN}}(F_{X \rightarrow Y}, D_Y) + L_{\text{GAN}}(G_{Y \rightarrow X}, D_X) + \lambda_{\text{CCL}} L_{\text{CCL}}(F_{X \rightarrow Y}, G_{Y \rightarrow X}), \quad (5)$$

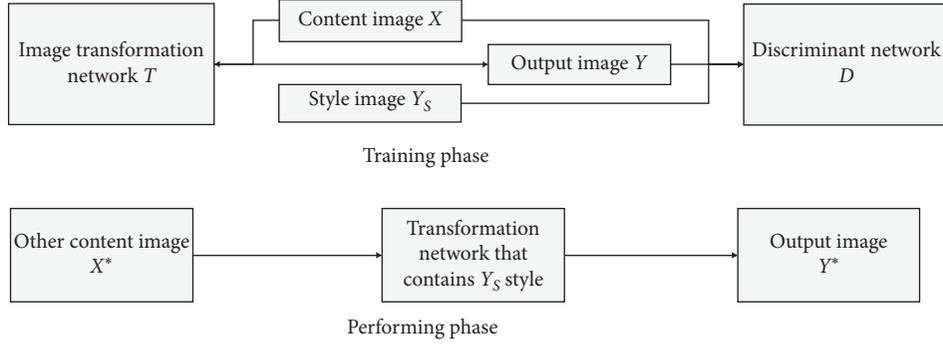


FIGURE 1: Image painting art style rendering process.

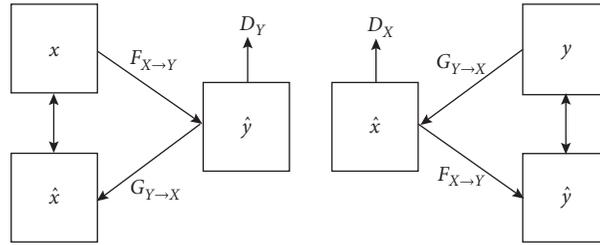


FIGURE 2: CycleGAN backward consistency and backward consistency.

where λ_{CCL} is the weight of adjusting the loss L_{CCL} in the objective function.

3.2. Image Conversion Network. The image style conversion network T improves on the network proposed in [22], and the image style conversion network structure is shown in Figure 3. The overall structure follows its deep residual network (ResNet), which has three convolutional layers and five residual blocks, as shown in Figure 3(a). Except that the output layer uses a scaled Tanh function to ensure that the output pixels are in range of $[0, 255]$, all other nonresidual convolutional layers use ReLU activation functions. Since no pooling layer is used and stride convolution or microstride convolution is used for upsampling and downsampling, this operation not only reduces the parameters, but also maintains a large field of view and avoids excessive deformation of image objects [23].

Considering that each nonresidual convolutional layer in the network is followed by a spatial batch normalization to achieve accelerated convergence and the study in [24] showing that instance regularization is better than batch normalization, which can significantly improve performance, this paper uses instance normalization instead of batch normalization, as shown in Figure 3(b). It can be seen that the instance normalization is actually to modify the batches to 1, and the normalization is used on a single image instead of a batch of images. Therefore, the network improved by the instance normalization performs better in the test phase of painting art style rendering.

3.3. Discriminator Network. The discriminant network D is a multilayer convolutional neural network, as shown in Figure 4. Behind each hidden layer are added the batch

normalization and LeakyReLU linear activation functions. The first, fourth, sixth, and eighth layers are used to measure the perceptual adversarial loss between the generated image and the style image. The judgment network outputs a probability of the images, that is, whether the image is from the real data set (True) or generated by the style conversion network (Fake) [10].

4. Painting Art Style Rendering Based on the CycleGAN Model and Perceptual Confrontation Loss

4.1. Perceptual Adversarial Loss. Although the PAN model [25] has been proven to measure the difference between images from a high-dimensional visual perception level, how to extract effective feature differences through the hidden layer is still an open question. The key to the problem is how to minimize the difference between the generated image and the real image in each high-dimensional level [26]. For this reason, this paper combines the PAN model and the perceptual loss proposed in [27] and defines the perceptual adversarial loss to consist of content feature loss, style loss, and adversarial loss. In the N -layer discriminant network, the image features are regarded as feature maps of N dimensions. If the size of each layer of feature maps is $H_i \times W_i$, the size of the feature map is $C_i \times H_i \times W_i$, where C represents the number of feature maps. Since each grid position of the image can be regarded as an independent sample, it can capture the key features. Perceptual adversarial loss is the weighted sum of content loss and style loss. It can punish the difference between the generated image and the style image when the first, fourth, sixth, and eighth hidden layers of the network are dynamically updated, so that the

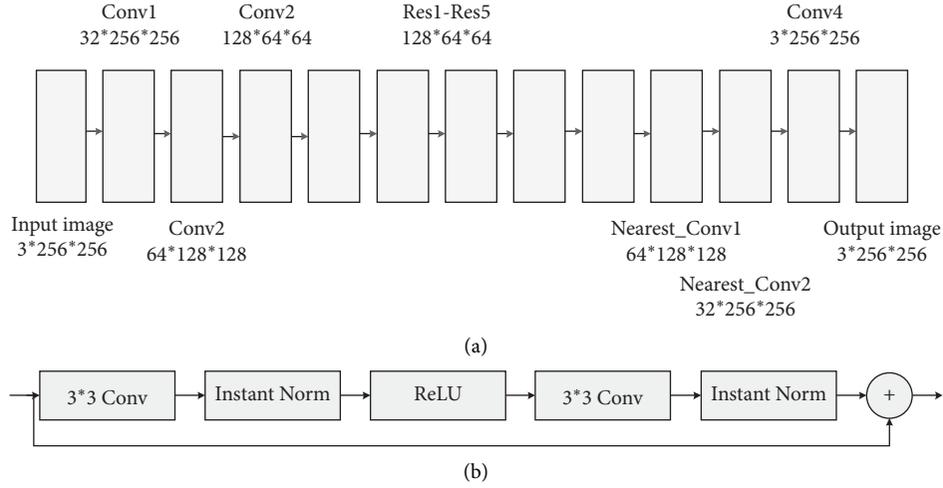


FIGURE 3: Image conversion network structure. (a) Artistic painting style rendering system model. (b) The residual structure used in the model.

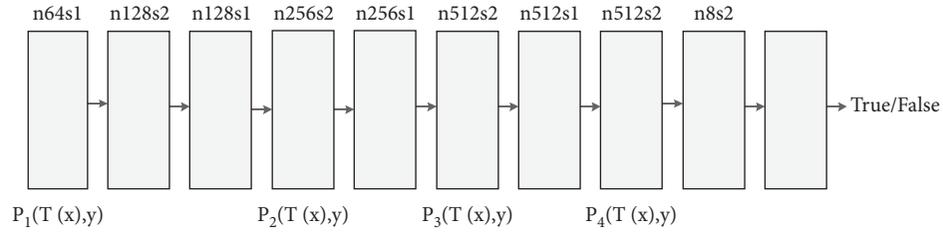


FIGURE 4: Discriminant network structure of the painting art style rendering system.

generated image has the best excellent content and style synthesis effect.

4.1.1. Content Loss Function. The content loss function P_i uses Manhattan distance to calculate the image space loss (L_2) of the hidden generated image Y and the real painting art style image X as follows:

$$P_i = \frac{1}{N} \|H_i(Y) - H_i(X)\|, \quad (6)$$

where H_i represents the L_2 value of the i -th hidden layer of the discrimination network.

Based on the above definition, multiple levels of content loss can be expressed as

$$L_{\text{content}}(X, Y) = \sum_{i=1}^N \lambda_i P_i(X, Y), \quad (7)$$

where λ_i represents the balance factor of discriminating the N hidden layers i of the network. By minimizing the perceptual loss function L_{content} , the generated image and the content image have a similar content structure.

4.1.2. Style Loss Function. Considering that the style loss function is used to penalize the deviation of the output image in style, including color and texture, this paper uses the style reconstruction method [28], which is obtained by the distance

between the output image and the style image Gram matrix. Set $\varphi_i(x)$ as the feature map of the i -th hidden layer, so that the shape of $\varphi_i(x)$ is $C_i \times (H_i \times W_i)$, and the style loss value of the feature map of the i -th layer of the discriminant network can be expressed as

$$\text{Gram}_i(Y) = \frac{\varphi_i(Y) * \varphi_i(Y)^T}{C_i \times H_i \times W_i}, \quad (8)$$

$$G_i(Y_s, Y) = \|\text{Gram}_i(Y) - \text{Gram}_i(Y_s)\|.$$

In order to represent the style reconstruction from multiple levels, this paper defines $G_i(Y_s, Y)$ as a set of losses (sum of losses for each layer) as

$$L_{\text{style}}(Y_s, Y) = \sum_{i=1}^n G_i(Y_s, Y). \quad (9)$$

4.1.3. Perceptual Adversarial Loss Function. The overall perceptual loss consists of a combination of content loss and style loss as a linear function, which can be defined as follows:

$$L_{\text{style}}(Y_s, Y) = \sum_{i=1}^n G_i(Y_s, Y), \quad (10)$$

where λ_c, λ_s are the weight parameters set based on human experience. The style conversion network T and the

discriminant network D are alternately optimized based on the overall perceptual loss value. The alternate optimization between such two network models is based on the method of PAN [29] model to realize the min-max confrontation. For the generated image Y , content image X , and style image Y_s , the loss function of network T and the loss function of network D are defined as follows:

$$\begin{aligned} L_T &= \ln(1 - D(T(x))) + L_{\text{perc}}, \\ L_D &= -\ln(D(y)) - \ln(1 - D(T(x))) + [m - L_{\text{perc}}]^+. \end{aligned} \quad (11)$$

We set a positive boundary value m in equation (11). Because the positive boundary value m can make the third term of L_D achieve gradient descent, we minimize L_T through the parameters of the network T , which can maximize the second and third terms of L_D at the same time. When L_T is less than m , the loss function L_D will update the discriminant network to a new high-dimensional level to calculate the remaining difference. Therefore, by perceiving against loss, the diversified differences between the generated image and the style image can be continuously perceived and explored.

4.2. Painting Art Segmentation Rendering System. The generator of the traditional CycleGAN model adopts ResNet [21]. The generator is a fully convolutional connection type [30] and consists of an encoder, a converter, and a decoder. On the basis of the perceptual loss error, this paper improves the traditional CycleGAN network and uses CycleGAN with the DensNet [31] generator to achieve the style rendering of the image, as shown in Figure 5.

4.2.1. Encoder. The CNN is used to extract features from the input image. The number of filters in the first convolutional layer (Conv_Layer_1) is set to 64. The output of the first convolutional layer is sent to the second convolutional layer (Conv_Layer_2) to continue to extract features, and the number of filters in the second convolutional layer is set to 128. The connection and data transmission method of the second convolutional layer and the third convolutional layer (Conv_Layer_3) is similar to that of the first and second convolutional layers. The number of filters in the third convolutional layer is set to 256. The image size input to the encoder is 256×256 , and the encoder extracts 256 feature vectors with a size of 64×64 .

4.2.2. Converter. The purpose of converter is to combine the different extracted features and use these features to determine how to convert the feature vector of the image from the X domain (image style template) to the feature vector of the Y domain (generated image). The original CycleGAN network converter uses a 6-layer ResNet block to convert the feature vector, as shown in the following equation:

$$f_l = [H_l(f_{l-1})] + [f_{l-1}], \quad (12)$$

where H_l is a nonlinear transformation function. f_{l-1} and f_l are the input features and style conversion output features of the first layer as a ResNet block, including batch normalization layer [25], convolutional layer, and ReLU layer [32]. The converter designed in this paper uses DenseNet model to replace the traditional ResNet model. DenseNet model can reduce the disappearance of the gradient, enhance the feature transfer ability, and reduce the number of parameters to a certain extent. DenseNet model connects modules to each other, which improves the coupling capability of information flow between modules. The l -th module of the DenseNet model accepts the feature mapping from the previous module, as shown in the following equation:

$$f_l = H_l([f_0, f_1, \dots, f_{l-1}]), \quad (13)$$

where $[f_0, f_1, \dots, f_{l-1}]$ is formed by the series of generation features of 0, 1, \dots , $l-1$ layers. This paper integrates the DenseNet module in the converter to reduce model parameters while avoiding overfitting and reducing the amount of calculation.

4.2.3. Decoder. The decoder process is opposite to that of the encoder, using a three-level deconvolution layer (Deconv_Layer) [33] to restore the low-level features of the image from the feature vector step by step, until the image is generated. The detailed internal composition of encoder, converter, and decoder is shown in Table 1.

4.2.4. Discriminator. The discriminator of the improved network adopts the PatchGAN [34] classifier. In this model, we input the image to the discriminator, which discriminates whether the image is the original image or the image generated by the generator. The image input to the discriminator is divided into multiple 70×70 image blocks. In the discriminator's discrimination process, the network convolves the input image layer by layer, discriminates each image block through the one-dimensional output convolutional layer, and takes the average of the judgment results of all image blocks as the image judgment result.

5. Simulation Experiment and Result Analysis

5.1. Experiment Initialization. The experimental environment used in this paper is as follows: CPU: Intel(R) Core(TM) i7-9700K@3.20 GHZ, 16 GB memory, operating system: 64 bit Windows 10, deep learning framework: Pytorch 1.10. This open source framework is mainly written in Python and can be used across platforms to implement various common CNN models and GAN models.

In order to verify the feasibility and effectiveness, the following model training and testing data sets are constructed for the perceptual adversarial loss error and the improved CycleGAN + PAN model. In the network training process, 983 oil painting images, 933 sketch painting images, and 1000 260×260 RGB ink wash paintings were selected as the image painting art style rendering data set. The training data used 1142 architectural images. There is no designated corresponding pairing relationship between different types

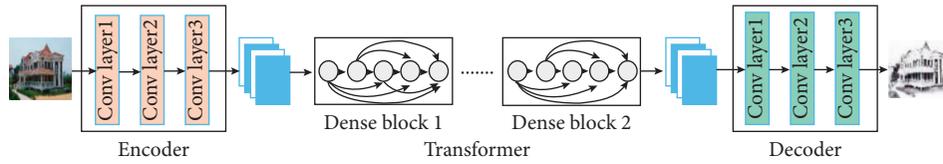


FIGURE 5: Generation network structure of the painting art style rendering system.

TABLE 1: Internal structure of the generation network of the painting art style rendering system.

Modules	Descriptions in detail			
Encoder	Convolutional kernel	Ways of network connecting	Number of filters	Stride of operations
1	7 * 7	Conv, BN, ReLU	64	1
2	5 * 5	Conv, BN, ReLU	128	1
3	3 * 3	Conv, BN, ReLU	256	2
Transformer	Modules			
1	Dense block 1			
...			
5	Dense block 5			
Decoder	Convolutional kernel	Ways of network connecting	Number of filters	Stride of operations
1	3 * 3	De-conv, BN, ReLU	128	1/2
2	3 * 3	De-conv, BN, ReLU	64	1/2
3	7 * 7	De-conv, BN, ReLU	3	1

of data sets during the training process. As shown in Figure 6, they are CycleGAN, CycleGAN + PAN, and the image style rendering algorithm based on the improved CycleGAN + PAN that is proposed in this paper, Ground-truth \rightarrow Oil painting, Ground-truth \rightarrow Sketch painting, and Ground-truth \rightarrow Ink wash painting for three styles of rendering model training.

5.2. Qualitative Results Analysis. In order to verify the feasibility and effectiveness of the rendering model proposed in this paper, the objective functions of CycleGAN, CycleGAN + PAN, and the improved CycleGAN + PAN are used in the experiment to perform image painting art style rendering experiments. The experimental results are shown in Figure 7. The first column in Figure 7 is the original image, the second column is the style image, and the third, fourth, and fifth columns are, respectively, experimental results of image style rendering for the CycleGAN model, CycleGAN + PAN model, and the improved CycleGAN + PAN model.

The first row, the second row, and the third row are experiments of three comparative models using oil painting, ink painting, and sketch painting as rendering styles. From the comparison of experimental results, it can be found that, under the same number of iterations, the proposed model can achieve painting art style rendering faster and achieve more realistic style rendering effects. Under the same experimental configuration and settings, this paper compares the generation results after 50,000 iterations. Table 2 shows the image style rendering results of CycleGAN, CycleGAN + PAN, and the proposed model using SSIM and PSNR image quality evaluation indicators for three different styles' rendering.

It can be seen from Table 2 that, in the results of the oil painting style rendering experiment, the PSNR of CycleGAN + PAN model is slightly higher than the proposed model. However, in the experimental results of the style rendering of ink and drawing styles, the SSIM and PSNR of the improved painting art style rendering of the proposed model are higher than other models.

5.3. Quantitative Results Analysis. CycleGAN introduces a cycle consistency loss function on the basis of traditional GAN unidirectional mapping and avoids model collapse to a certain extent through bidirectional mapping. The improved CycleGAN + PAN model proposed in this paper introduces the same mapping loss and perceptual loss function. Compared with CycleGAN + PAN model, the image style rendering process is more stable and the model is not easy to collapse. Figure 8 shows the comparative experimental results of some detailed painting art rendering between CycleGAN + PAN and the improved CycleGAN + PAN model proposed in this paper. Under the same number of iterations, the results of image style rendering are compared for two hard situations, which are marked as red box in Figure 8.

It can be seen from Figure 8 that CycleGAN + PAN uses two-way mapping to reduce the probability of model collapse, but instability sometimes occurs during training. As shown in Figure 8, there are black blocks in the image style rendering with the CycleGAN + PAN training model, and the style is lost to a certain extent. The third column shows the style rendering image generated by the improved CycleGAN + PAN. The result is that the image rendering style is well preserved, and there is no model crash. Therefore, the improved CycleGAN + PAN model proposed



FIGURE 6: An example display of the experimental data set of painting art style rendering. (a) Ground-truth. (b) Oil painting. (c) Sketch painting. (d) Ink wash painting.



FIGURE 7: Experimental results of painting art style rendering. (a) Oil painting-based painting art style rendering. (b) Sketch painting-based painting art style rendering. (c) Ink wash painting-based painting art style rendering.

TABLE 2: Quality evaluation of the painting art style rendering experiment results.

Different styles	Indexes	CycleGAN	CycleGAN + PAN	Proposed method
Oil painting	SSIM	0.5451	0.7216	0.7456
	PSNR	18.6212	20.4515	20.4425
Sketch painting	SSIM	0.6315	0.6418	0.6987
	PSNR	18.5187	18.9251	20.2147
Ink wash painting	SSIM	0.6542	0.6781	0.7121
	PSNR	17.8154	18.1214	19.0241

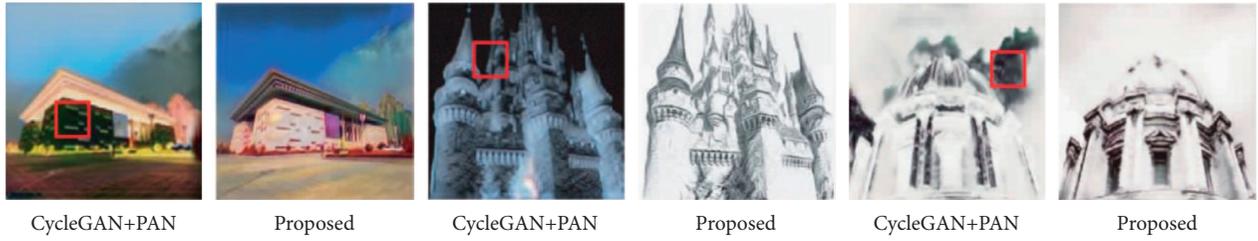


FIGURE 8: Effect of the abnormal situation of painting art style rendering between the original CycleGAN + PAN model and the improved CycleGAN + PAN model proposed in this paper.

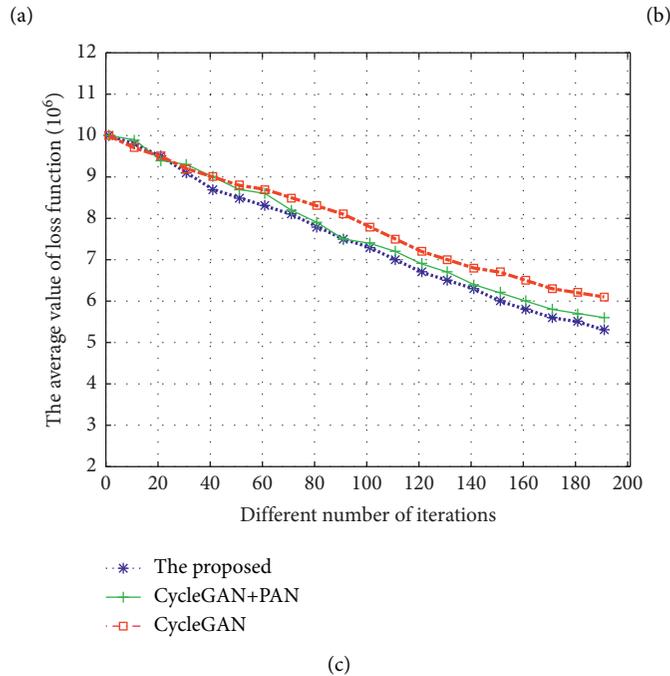
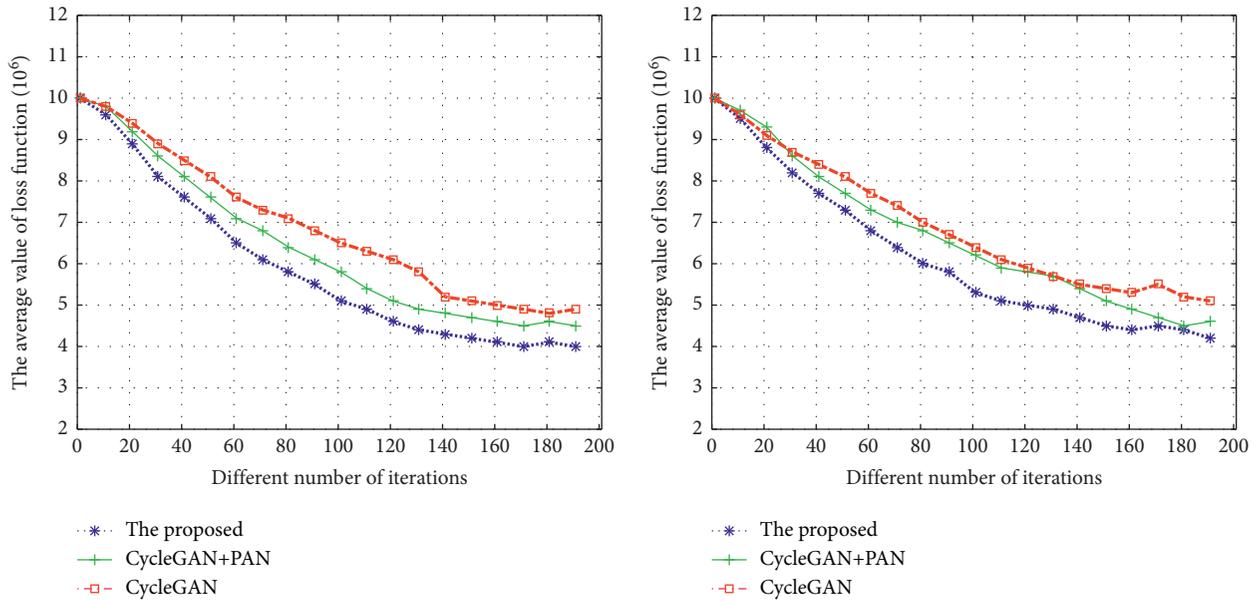


FIGURE 9: The comparison results of the loss function of the improved model in this article, the classic CycleGAN + PAN model, and the CycleGAN model. (a) 256×256 . (b) 512×512 . (c) 1024×1024 .

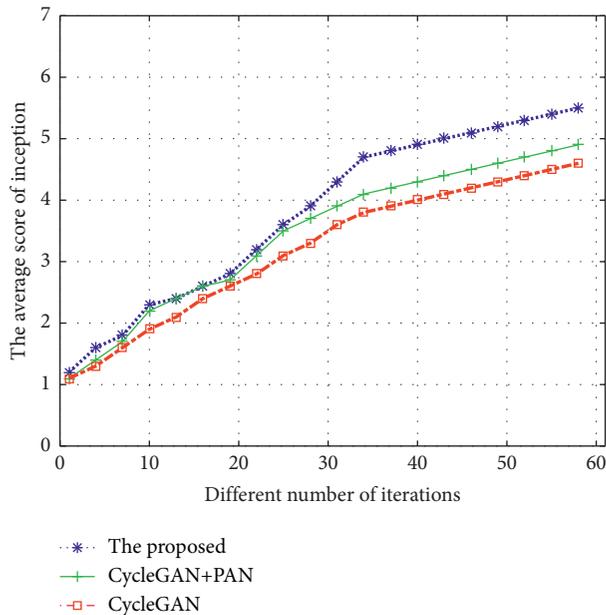


FIGURE 10: Comparison of the efficiency of the three comparison models in the rendering of artistic style of painting.

in the paper has good training conditions and is not easy to collapse under the same number of iterations.

The purpose of considering the painting art style rendering method is to achieve style loss and content loss, that is, to minimize equation (10). Therefore, the method proposed in this paper can be compared with the original CycleGAN + PAN [21] model in terms of measuring the extent of the successful reduction of the loss function. This paper uses the three compared methods to train 50, 100, and 200 images to record the value of equation (10) for each group of training. The loss function values of the three different sizes of images are shown in Figure 9. It can be seen from Figure 9 that, no matter for low-resolution or high-resolution images, the perceptual adversarial network can more effectively minimize the loss function.

The method proposed in this paper improves the convergence speed more than the original CycleGAN model. As shown in Figure 10, CycleGAN model, CycleGAN + PAN model, and improved CycleGAN + PAN model are trained on the same data set, and the inception score is used to measure the quality and variety of style rendering images. In the case of the same number of iterations, the higher the evaluation score is, the better the quality of the style rendering image is and the faster the convergence speed is obtained. In the experimental results of the three models, the improved CycleGAN + PAN model can converge faster and obtain higher evaluation scores. Through the convergence speed comparison experiment, we can see that the improved CycleGAN + PAN model proposed in this paper has a faster convergence speed and better stylized image quality.

6. Conclusion

In this paper, we select the characteristics of the PAN model and CycleGAN model and combine the advantages of such

two models to construct the image painting art style rendering system. The discriminator updates parameters to explore the difference between generated images and style image, and the image conversion network uses the discriminant network to proceed before feed training until the difference is minimized to the optimal image conversion model. Such two models are alternately updated according to the perceptual loss. Experimental results have shown that, compared with the original CycleGAN network with the ResNet module, this paper adopts the improved CycleGAN + PAN image style rendering system based on perceptual confrontation loss. The style rendering effect is better, and the learned style can be rendered to new painting art style image to obtain a clearer performance effect. In future research, further optimization of the network structure and loss function will be considered in our researches. In terms of color and texture, the structural similarity and semantic consistency of content maps and style maps will be enhanced for novel combination of network structure and loss functions.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

- [1] P. Andreini, S. Bonechi, M. Bianchini, M. Alessandro, and S. Franco, "Image generation by GAN and style transfer for agar plate image segmentation," *Computer Methods and Programs in Biomedicine*, vol. 184, Article ID 105268, 2020.
- [2] C. T. Lin, S. W. Huang, Y. Y. Wu, and H. L. Shang, "GAN-based day-to-night image style transfer for nighttime vehicle detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 951–963, 2020.
- [3] S. Yang, Z. Wang, L. Jiaying, Z. Gao, N. Xu, and J. Liu, "Controllable artistic text style transfer via shape-matching gan," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4442–4451, IEEE, Seoul, South Korea, October 2019.
- [4] Y. Jing, Y. Yang, Z. Feng, Y. Jingwen, Y. Yizhou, and S. Mingli, "Neural style transfer: a review," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 11, pp. 3365–3385, 2019.
- [5] X. Chen, C. Xu, X. Yang, S. Li, and T. Dacheng, "Gated-gan: adversarial gated networks for multi-collection style transfer," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 546–560, 2018.
- [6] L. Zhang, Y. Ji, and X. Lin, "Style transfer for anime sketches with enhanced residual u-net and auxiliary classifier gan," in *Proceedings of the 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 506–511, IEEE, Nanjing, China, December 2017.
- [7] X. Deng, R. Yang, M. Xu, and L. D. Pier, "Wavelet domain style transfer for an effective perception-distortion tradeoff in

- single image super-resolution,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3076–3085, IEEE, Seoul, South Korea, October 2019.
- [8] X. Deng, “Enhancing image quality via style transfer for single image super-resolution,” *IEEE Signal Processing Letters*, vol. 25, no. 4, pp. 571–575, 2018.
 - [9] Y. Liu, “Improved generative adversarial network and its application in image oil painting style transfer,” *Image and Vision Computing*, vol. 105, Article ID 104087, 2021.
 - [10] H. Fang, W. Deng, Y. Zhong, and H. Jiani, “Triple-GAN: progressive face aging with triple translation loss,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 804–805, IEEE, Seattle, WA, USA, July 2020.
 - [11] J. Ye, Y. Ji, X. Wang, G. Xin, and M. Song, “Data-free knowledge amalgamation via group-stack dual-gan,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, August 2020, Article ID 12516.
 - [12] H. Emami, M. M. Aliabadi, M. Dong, and B. C. Ratna, “Spagan: spatial attention gan for image-to-image translation,” *IEEE Transactions on Multimedia*, vol. 23, pp. 391–401, 2020.
 - [13] H. Liu, P. N. Micheline, D. Zhu, and G. Artsy, “A style transfer system with improved quality, diversity and performance,” in *Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 79–84, IEEE, Beijing, China, November 2018.
 - [14] A. A. Atapour and T. P. Breckon, “Real-time monocular depth estimation using synthetic data with domain adaptation via image style transfer,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2800–2810, IEEE, Salt Lake City, UT, USA, December 2018.
 - [15] R. Li, C. H. Wu, S. Liu, W. Jue, and W. Guangfu, “SDP-GAN: saliency detail preservation generative adversarial networks for high perceptual quality style transfer,” *IEEE Transactions on Image Processing*, vol. 30, pp. 374–385, 2020.
 - [16] J. Svoboda, A. Anooosheh, C. Osendorfer, and M. Jonathan, “Two-stage peer-regularized feature recombination for arbitrary image style transfer,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, August 2020, Article ID 13816.
 - [17] Y. Zhang, Y. Zhang, and W. Cai, “Separating style and content for generalized style transfer,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8447–8455, IEEE, Salt Lake City, UT, USA, June 2018.
 - [18] P. Li and M. Yang, “Semantic GAN: application for cross-domain image style transfer,” in *Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 910–915, IEEE, Shanghai, China, August 2019.
 - [19] Z. Ma, J. Li, N. Wang, and G. Xinbo, “Semantic-related image style transfer with dual-consistency loss,” *Neurocomputing*, vol. 406, pp. 135–149, 2020.
 - [20] T. Wang, W. Q. Toh, H. Zhang, and S. Xiuchao, “RoboCo-DRAW: robotic avatar drawing with GAN-based style transfer and time-efficient path optimization,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, New York, NY, USA, February 2020.
 - [21] C. Liu, X. Chang, and Y. D. Shen, “Unity style transfer for person re-identification,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6887–6896, IEEE, Seattle, WA, USA, August 2020.
 - [22] R. Kips, P. Gori, M. Perrot, and B. Isabelle, “Ca-gan: weakly supervised color aware gan for controllable makeup transfer,” in *Proceedings of the European Conference on Computer Vision*, pp. 280–296, Springer, Glasgow, UK, August 2020.
 - [23] P. Wang, Y. Li, and N. Vasconcelos, “Rethinking and improving the robustness of image style transfer,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 124–133, New Orleans, LA, USA, November 2021.
 - [24] Y. Zhou, K. Yu, M. Wang et al., “Speckle noise reduction for OCT images based on image style transfer and conditional GAN,” *IEEE Journal of Biomedical and Health Informatics*, 2021.
 - [25] Y. Chong, C. Peng, J. Zhang, and P. Shaoming, “Style transfer for unsupervised domain-adaptive person re-identification,” *Neurocomputing*, vol. 422, pp. 314–321, 2021.
 - [26] Z. Yang, Z. Hu, C. Dyer, and P. X. Eric, “Unsupervised text style transfer using language models as discriminators,” in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 7298–7309, Montréal, Canada, December 2018.
 - [27] S. Yang, J. Liu, W. Wang, and G. Zongming, “TET-GAN: text effects transfer via stylization and destylization,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, Honolulu, HI, USA, February 2019.
 - [28] Y. Men, Y. Mao, Y. Jiang, Y. M. Wei, and L. Zhouhui, “Controllable person image synthesis with attribute-decomposed gan,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5084–5093, IEEE, Seattle, WA, USA, August 2020.
 - [29] A. Zhu, X. Lu, X. Bai, U. Seiichi, K. I. Brian, and X. Shengwu, “Few-shot text style transfer via deep feature similarity,” *IEEE Transactions on Image Processing*, vol. 29, pp. 6932–6946, 2020.
 - [30] Y. Xu, Y. Li, and B. S. Shin, “Medical image processing with contextual style transfer,” *Human-centric Computing and Information Sciences*, vol. 10, no. 1, pp. 1–16, 2020.
 - [31] Z. Hu, J. Jia, B. Liu, F. Jianlong, and B. Yaohua, “Aesthetic-aware image style transfer,” in *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 3320–3329, Seattle, WA, USA, October 2020.
 - [32] M. Lu, H. Zhao, A. Yao, C. Yurong, and Z. Li, “Decoder network over lightweight reconstructed feature for fast semantic style transfer,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2469–2477, IEEE, Venice, Italy, December 2017.
 - [33] S. Zhou, M. Ke, and P. Luo, “Multi-camera transfer GAN for person re-identification,” *Journal of Visual Communication and Image Representation*, vol. 59, pp. 393–400, 2019.
 - [34] Y. Huang, M. He, L. Jin, and W. Yongpan, “RD-GAN: few/zero-shot Chinese character style transfer via radical decomposition and rendering,” in *Proceedings of the European Conference on Computer Vision*, pp. 156–172, Springer, Glasgow, UK, August 2020.