

Research Article

Digital Image Art Style Transfer Algorithm Based on CycleGAN

Xuhui Fu 

Culture and Art Management of Hunan University, Korea, Guangzhou City 62399, Republic of Korea

Correspondence should be addressed to Xuhui Fu; yun8876599@my.honam.ac.kr

Received 21 October 2021; Revised 16 November 2021; Accepted 25 November 2021; Published 13 January 2022

Academic Editor: Bai Yuan Ding

Copyright © 2022 Xuhui Fu. 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.

With the continuous development and popularization of artificial intelligence technology in recent years, the field of deep learning has also developed relatively rapidly. The application of deep learning technology has attracted attention in image detection, image recognition, image recoloring, and image artistic style transfer. Some image art style transfer techniques with deep learning as the core are also widely used. This article intends to create an image art style transfer algorithm to quickly realize the image art style transfer based on the generation of confrontation network. The principle of generating a confrontation network is mainly to change the traditional deconvolution operation, by adjusting the image size and then convolving, using the content encoder and style encoder to encode the content and style of the selected image, and by extracting the content and style features. In order to enhance the effect of image artistic style transfer, the image is recognized by using a multi-scale discriminator. The experimental results show that this algorithm is effective and has great application and promotion value.

1. Introduction

The specific characteristics of artistic schools are mainly artistic styles, and different styles also present different cultural connotations and backgrounds. Image art style transfer is mainly the processing of digital images, which refers to the method of forming new image features by fusing the features of different images with each other. There are more and more achievements in the field of deep learning. For example, in 2015, scholar Gatys proposed a method of image art style transfer, through image reorganization, segmentation, and other methods, to generate a confrontation network model. It has also become particularly applicable in image processing, which has attracted the attention of many scholars. The main principle is to improve the performance of neural networks through adversarial training. In order to ensure the effectiveness and operability of the network, the applicability of the model is ensured by adjusting the introduction mechanism of the generative confrontation network and adjusting the frame result. This paper proposes a new artistic style based on the generative confrontation network model, through the extraction of image features and the transfer of image artistic styles. This article focuses on how to finally achieve

the artistic visual effects that need to be presented based on the generative confrontation network model. Some problems that can be improved are discovered, and then verified by experiments, practical solutions are proposed, which have certain practical application value and important research significance.

2. Related Work

Deep neural networks contain one or more hidden layers of neural network structures, which have good generalization capabilities for complex problems. In theory, deep learning networks can fit arbitrary complex functions and show a powerful ability to learn sample features. In deep learning, CNN is the most commonly used deep neural network for processing computer vision tasks. By sharing the convolution kernel, the CNN can process high-dimensional data, and as long as the training parameters are set, the features in the image can be extracted through layered convolution. Therefore, deep neural networks based on CNN are good at learning and extracting the features and special information contained in images, such as texture features, shape features, boundary features, grayscale features, topological features, and relational structures. The extraction and utilization of

these features or information make the application of image processing further enriched and expanded.

Different experts and scholars have carried out different researches on image art style transfer processing. Among them, Sun et al. [1] proposed a CNN architecture with two paths, by extracting the object features and textures of the image, respectively. Features realize image style recognition and output image category. Zhao et al. proposed a [2] high-definition image deep learning training algorithm. The method is mainly divided into three steps: (1) break the traditional method of extracting image features, by extracting the image features of high- and low-resolution images, which can greatly save training time; (2) establish a high-resolution deep learning model based on the dictionary model; (3) transform low-resolution images into high-resolution images. Lu et al. [3–5] believed that the current deep learning model can no longer adapt to the aggregation of multiple images, and proposed a multi-image training model based on an improved neural network. Its framework is mainly to integrate the feature extraction and aggregation algorithms so that the image can extract multiple image features. Castillo et al. proposed constructing a model of artistic target transfer, its principle is mainly to carry out artistic transfer and fusion of a single image object. Regarding the artistic style transfer of dynamic video, Ruder et al. improved on the basis of Gatys, Gatys mainly transfers the artistic style of still images, and Ruder proposes to extract the artistic style of the video image, and extract and transfer the artistic style of the entire video. With the continuous extension and development of image processing methods and technologies in recent years, some creative thinking and methods have also been continuously developed. This paper proposes a generative confrontation network model, which encodes the style and content of the image based on the residual network structure and the encoding and decoding structure. The most important part of network decoding is the generative network, which uses the encoded information to generate the image of artistic style transfer, discriminate the network, make the image effect of the generated network better, and the image more vivid and lifelike, thereby improving the transfer effect of the entire image art style [6–10].

3. Related Theoretical Methods

3.1. The Algorithm Flow of Generating a Confrontation Network

3.1.1. Algorithm Description. GAN is a powerful generative model. The model is mainly composed of two parts; First, the generative model, which is mainly to generate natural and real images as much as possible; second is the discrimination model, which mainly judges the authenticity of the image [11].

3.1.2. GAN Algorithm Flow. The ultimate goal of GAN is to realize the transfer of artistic style by judging model D and generating model G. G and D are generally nonlinear

mapping functions, usually in the form of a network structure such as a multilayer perceptron or a convolutional neural network change. Assuming that $P_g(z) = \{z(1), \dots, z(n)\}$ is a given random noise sample set that obeys a simple distribution, $P_{data}(x) = \{x(1), \dots, x(n)\}$ is the real data sample set, G is the generating network, and D is the discriminant network (Algorithm 1) [8].

3.1.3. CycleGAN Algorithm. CycleGAN can transform two different styles of images, such as the mutual conversion of photos and oil paintings, day and night, spring, summer, autumn, and winter. Assume that X and Y are data sample sets in two different fields, G is a generator generated from X field to Y field, F is a generator generated from Y field to X field, it is to identify the X field data (real X field sample). The true and false discriminator of the X field data is generated by F, and F is a discriminator for authenticating the Y field data (real Y field samples and Y field data generated by G), which is a cycle consistent loss. The CycleGAN algorithm is described as follows (Algorithm 2) [9].

3.2. Image Art Style Transfer Method. The essence of the image to image conversion task is to learn the mapping between X and Y image domains: $G_{XY}: X \rightarrow Y$ and $G_{YX}: Y \rightarrow X$. The goal of the unsupervised translation is to restore this mapping using unpaired samples in the original data. This method combines the instance information and adds a set of instance attribute information to the original image x to improve the image to image translation effect. Therefore, this method can be understood as learning $X \times A$ and $Y \times A$ for the joint mapping between B, each network is designed to simultaneously encode the image and the corresponding instance mask. The method is described in detail as follows:

CycleGAN is selected as the basic structure, and the periodic consistency loss is used to encourage $G_{YX}(G_{YX}(x)) \approx x$ and $G_{XY}(G_{XY}(y)) \approx y$. As shown in Figure 1, it is composed of two generators g and F , and two discriminators D_Y and D_X . It is assumed that the given source image data set and target image data set are $\{x_i\}_{i=1}^N \in X$ and $\{y_j\}_{j=1}^M \in Y$, N and M represent the data set size of the source image and the target image, respectively, x_i represents the first image in the source image, and y_j represents the j^{th} image in the target image.

The adversary discriminator D_Y is responsible for distinguishing the original image $\{y\}$ from the generated image $\{G(x)\}$, and the adversary discriminator D_X is also responsible for distinguishing the original image $\{x\}$ from the generated image $\{F(y)\}$. This structure is circular, that is, converting from one domain to another and then returning. Among them, generators G and F and discriminators D_Y and D_X follow the structure in Figure 1

For the task of this paper, the structure is also maintained, and the whole includes two generators G_{XY} and G_{YX} , as well as the corresponding adversary discriminators D_Y and D_X . First train the generator $G_{XY}: X \times A \rightarrow Y$, which simultaneously encodes and converts the original data (x, a) to the target domain data y' . The generator first inputs the original photo and the corresponding segmentation mask,

```

BEGIN
FOR training iterations DO:
Extract m noise samples  $[z^{(1)}, \dots, z^{(m)}]$  from the noise data  $p_g(z)$ .
Extract m noise samples  $[x^{(1)}, \dots, x^{(m)}]$  from the real data  $p_{data}(x)$ .
Optimize the parameters of the discriminant network through stochastic gradient ascent.
 $\nabla_{\theta_d} (1/m) \sum_{i=1}^m [\log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))]$ .
Extract m noise samples  $[z^{(1)}, \dots, z^{(m)}]$  from the noise data  $p_g(z)$ .
Generate network parameters through stochastic gradient descent optimization.  $\nabla_{\theta_d} (1/m) \sum_{i=1}^m [\log(1 - D(G(z^{(i)})))]$ 

```

ALGORITHM 1: GAN algorithm.

```

BEGIN
FOR training iterations DO:
Extract n sample data from the X field  $\{x^{(1)}, \dots, x^{(n)}\}$ .
Extract m sample data from the Y field  $\{y^{(1)}, \dots, y^{(m)}\}$ .
 $l_{real}^{(D)} = (1/n) \sum_{i=1}^n (D_x(x^{(i)}) - 1)^2 + (1/m) \sum_{i=1}^m (D_y(y^{(i)}) - 1)^2$ .
Calculate the discriminator loss of the generated data  $l_{fake}^{(D)} = (1/n) \sum_{i=1}^n (D_y(G(x^{(i)})) - 1)^2 + (1/m) \sum_{j=1}^m (D_x(F(y^{(j)})))^2$ .
Optimize the discriminator network parameters.
Calculate the generator loss of  $Y \rightarrow X$ 
 $l^F = (1/m) \sum_{i=1}^m (D_x(F(y^{(i)})) - 1)^2 + l_{cycle}^{Y \rightarrow X \rightarrow Y}$ .
Calculate the generator loss of  $X \rightarrow Y$ 
 $l^G = (1/n) \sum_{i=1}^n (D_y(G(x^{(i)})) - 1)^2 + l_{cycle}^{X \rightarrow Y \rightarrow X}$ .
Optimize generator network parameters.
END FOR.
END.

```

ALGORITHM 2: CycleGAN algorithm.

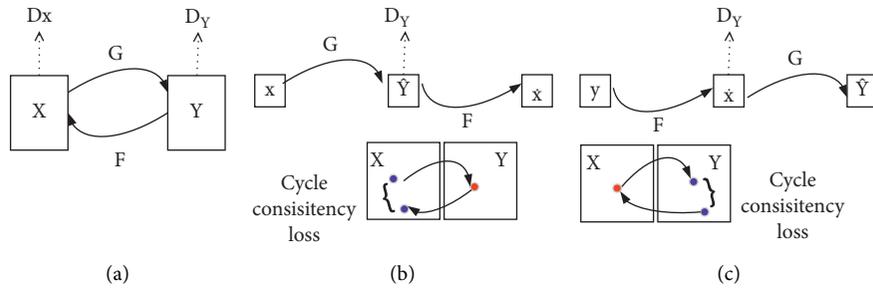


FIGURE 1: Periodic consistent generation countermeasure network model.

then uses the image feature extractor to extract features from the photo and mask, respectively, connects the corresponding image with the mask features, feeds them back to the image generator, and finally outputs the transformed image. The adversary discriminator $D_Y: Y \times B$ ("Y," "not Y"), where D_Y encodes y and b (or y' and b') at the same time and then determines whether the data is in the target domain $Y \times B$. The results are fed back to the generator or discriminator to be improved. The specific structure is shown in Figure 2.

For generator $G_{YX}: Y \times B \rightarrow X$, which simultaneously encodes and converts the original data (y', b') to the target domain data X . The generator first inputs the photo generated by G_{XY} and its corresponding segmentation mask, then uses the image feature extractor to extract features from

the photo and mask, respectively, connects the corresponding image with the mask features, feeds them back to the image generator, and finally outputs the transformed image. The adversary discriminator $D_X: X \times A$ ("x," "not X"), where D_X encodes x' and a (or x and a) at the same time, then distinguishes whether the data is in the target domain $X \times A$, and feeds back the results to the generator or discriminator to be promoted. The specific structure is shown in Figure 3.

Through the continuous game and learning using the generator and discriminator, the generator can finally generate an image with the style of the target area, while keeping the background and corresponding structure position completely consistent with the original image. For our task, input the image containing the target clothing and the

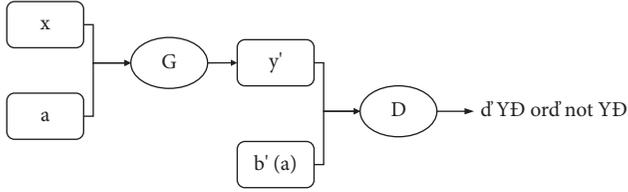


FIGURE 2: Generator and discriminator model 1.

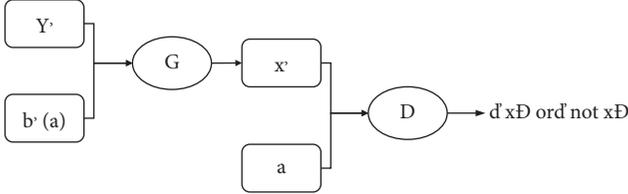


FIGURE 3: Generator and discriminator model 2 as a whole.

segmentation mask corresponding to the target clothing. After the whole model generation process, we can finally fix the clothing style migration task in the specific area of the target clothing, that is, complete the corresponding style migration of the target clothing while keeping the content and style of other areas unchanged. In addition, according to the target task requirements of this paper, the corresponding mask should be completely consistent before and after the image conversion, that is, the mask of the same image input generator should be the same as that in the discriminator in the whole process, which can be expressed as $a = b'$.

4. Image Art Style Transfer Based on Generative Confrontation Network

This article mainly improves the CycleGAN network to realize the artistic style transfer of images. Through the improvement of the CycleGAN network, the effect of image art style transfer is improved on the premise of reducing the calculation resources and calculation time occupied by the algorithm. Partial optimizations have been made in all aspects of training for research purposes. This chapter will sequentially introduce the image art style transfer network design based on the generative confrontation network, the design of the loss function, the training process and testing process of the network, the experimental environment, and the analysis of the experimental results [12–14].

4.1. Image Art Style Transfer Network Design Based on Generative Confrontation Network

4.1.1. Encoder Design

(1) *Content Coding.* The new content encoder in the network structure of this article adopts the autoencoder network model. The autoencoder uses a neural network architecture. In the self-encoder, the network part from the input layer to the coding layer is called the encoder, and the network part from the coding layer to the output layer is called the decoder [15].

The content encoder in the network structure of this article is mainly composed of the residual network. In 2016, He et al. suggested residual learning in the framework of ResNet, the input of the residual network is x , which is added to the output $F(x)$ of the function through the shortcut operation, and the output is $F(x)$ and x $F(x) + x$. The shortcut connection in the residual network increases the number of layers of the network, makes the extracted image features richer, and can effectively avoid the gradient disappearance and gradient explosion of the network [16]. The deep residual network can be effectively applied to the fields of image classification, image target detection, and image target localization. The content encoder in this article mainly uses an improved residual network, adding an instance normalization layer and a regular activation function after the convolutional layer. The schematic diagram of the residual structural unit in the content encoder is shown in Figure 4.

The network structure diagram of the content encoder is shown in Figure 5.

The specific process of the content encoder is described as follows:

- (1) The input picture is a 256×256 pixel three-channel image, which is down-sampled; the first layer uses a 7×7 size convolution kernel, the number of convolution kernels is 64, the sliding step is 1, and the filling size is 3, then the instance normalization process is performed, and finally, the ReLU activation process is performed.
- (2) The second and third layers both use 4×4 convolution kernels, the number of convolution kernels is 128 and 256, respectively, the sliding step size is 2 for both layers, and the filling size is 1, and then the instance is normalized, and finally the ReLU activation processing is carried out.
- (3) Then use 2 residual modules, there are two convolutional layers in the residual module, the number of convolution kernels in each convolutional layer is 256, the size of the convolution kernel is 3×3 , and the sliding step size is 1 for all, the padding size is 1, then the instance normalization process is performed, and finally the ReLU activation process is performed; specifics are as shown in Table 1.

(2) *Style Coding.* The new style encoder in the network structure of this paper adopts the variational autoencoder network model. Variational autoencoders require the hidden vectors generated by encoding to follow a Gaussian distribution. This restriction enables the encoder to understand the underlying laws of the training data and enables the encoder to learn the hidden variable model of the input data. The ordinary autoencoder learns a certain function through training data, and the variational autoencoder can learn the probability distribution of the parameters based on the training data. The style encoding of the image adopts a variational autoencoder to increase the randomness of style features [17].

The specific implementation method of the variational autoencoder is to change the output result of the encoder from one to two in the encoding stage, and the two vectors

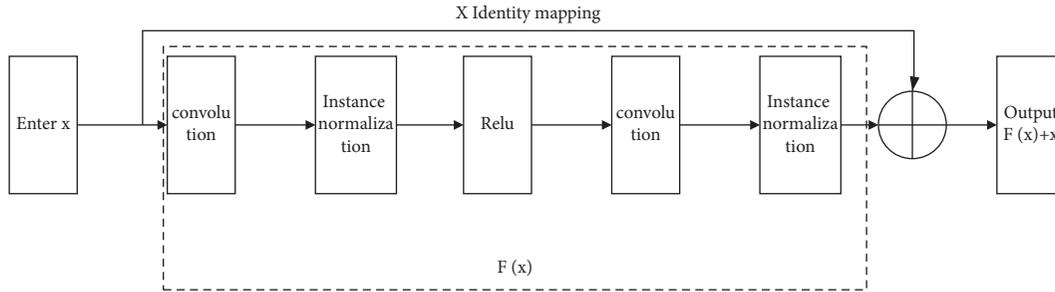


FIGURE 4: Residual structure unit in the content encoder.

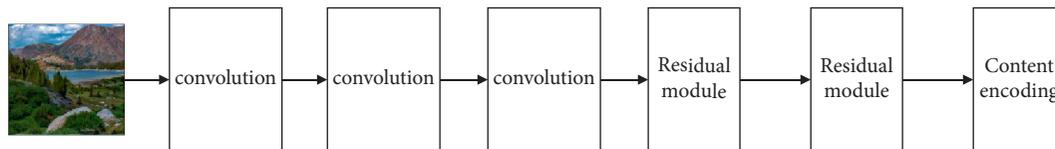


FIGURE 5: Content encoder structure diagram.

TABLE 1: Network information table of content encoder.

Operate	Kernel	Strides	Padding	Norm	Activation
Conv	(7,7,64)	1	3	Instance_Norm	ReLU
Conv	(4,4,128)	2	1	Instance_Norm	ReLU
Conv	(4,4,256)	2	1	Instance_Norm	ReLU
Resblock1	(3,3,256)	1	1	Instance_Norm	ReLU
Resblock2	(3,3,256)	1	1	Instance_Norm	ReLU

correspond to the mean vector and the standard deviation vector. and the hidden code vector is obtained by random sampling of the probability model [18–20].

The variational autoencoder can form a hidden vector of Gaussian distribution, randomly sample a hidden encoding vector from the Gaussian distribution as the style feature of the image, and then input the extracted style feature into the generation network. The network structure diagram of the style encoder is shown in Figure 6.

The specific process of the style encoder is described as follows:

- (1) The input picture is a 256×256 pixel three-channel image, which is down-sampled; the first layer uses a 7×7 convolution kernel, the number of convolution kernels is 64, the sliding step is 1, and the filling size is 3, then proceed to the ReLU activation reason.
- (2) The second to fifth layers all use 4×4 convolution kernels, the number of convolution kernels are 128, 256, 256, 256, respectively, the sliding step size is 2, the padding size is 1, and the ReLU activation processing is performed; then the average pooling operation is carried out.
- (3) The last layer uses a 1×1 size convolution kernel to realize the function of full connection. The number of convolution kernels is 8, the sliding step is 1, no padding, and no activation function; specifics are as shown in Table 2.

4.1.2. Generating Network Design. This article improves the generation network. In the process of implementing the image art style migration of the CycleGAN network, the generation network uses the U-net jump cascade network structure. The U-net performs multiple splicing in the decoding part and the encoding part, not only the previous decoding The result is used as input and the corresponding encoding part is also used as input for decoding so that the detailed information and high-level semantic information of the input image can be retained, but the U-net network structure occupies a large amount of memory, which increases the number of network calculations and the generated image. The result of artistic style transfer is rather vague [21–23]. In this paper, the generation network mainly uses the residual network structure, and at the same time, the improved generation networks G1 and G2 have the same network structure, but the purpose is different. The purpose of generating the network G1 is to receive the content and style features of the image, and then through the calculation of the intermediate hidden layer, so that the content of the image after the artistic style transfer remains unchanged, which improves the robustness of the network and makes the generated image more realistic. The generation network is a fully convolutional network, which uses residual structural units, up-sampling, and convolution operations to complete the artistic style transfer of the image. The residual structural unit used in the generation network is different from the residual structural unit used in the content encoder.

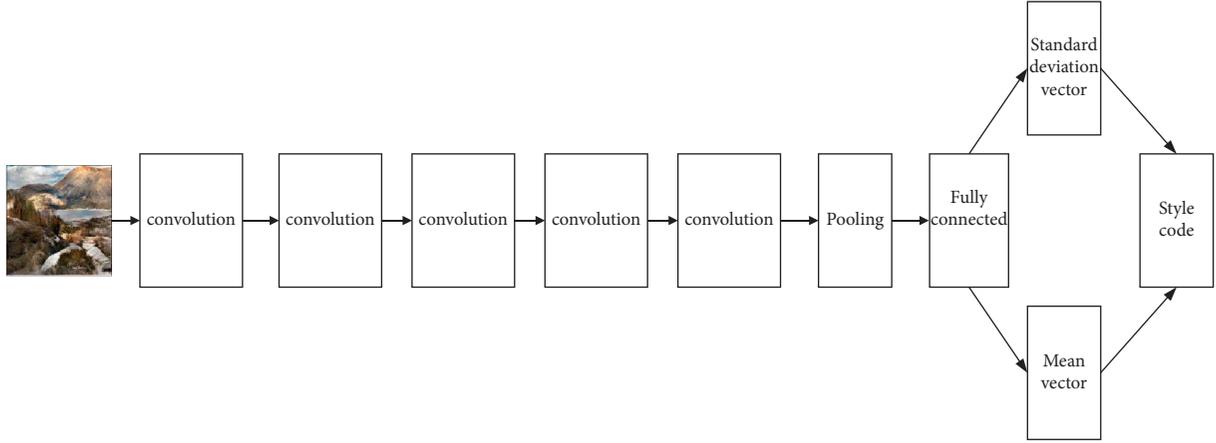


FIGURE 6: Structure diagram of style encoder.

TABLE 2: Style encoder network information table.

	Kernel	Strides	Padding	Activation
Conv	(7,7,64)	1	3	ReLU
Conv	(4,4,128)	2	1	ReLU
Conv	(4,4,256)	2	1	ReLU
Conv	(4,4,256)	2	1	ReLU
Conv	(4,4,256)	2	1	ReLU
Pooling	—	—	—	—
Fc	(1,1,8)	1	—	—

The normalization method of the residual structural unit in the network uses adaptive normalization. The adaptive normalization method can speed up the stylization of the image and make the generated image retain a strong style. In the generation network, the content information and style information is first input into the residual network at the same time, and after 4 residual module units, specifics are as shown in Figure 7.

The generative network is a convolutional neural network. The input is style features and content features, and the output is an image after artistic style transfer. The specific process is described as follows:

- (1) The input style coding and content coding first go through 4 residual network structures, in which each residual network structure uses a 3×3 size convolution kernel, the number of convolution kernels is 256, the sliding step size is 1, the padding size is 1, the normalization method uses adaptive normalization, and then performs the ReLU activation processing.
- (2) Use the nearest neighbor interpolation to double the size of the image, and then uses a 5×5 size convolution kernel. The number of convolution kernels are 128, 64, sliding step lengths are 1 for both, the padding size is 2, and the normalization method uses layer normalization, and then performs the ReLU activation processing.
- (3) The last layer uses a 7×7 convolution kernel, the number of convolution kernels is 3, the sliding step size is 1, and the padding size is 3. Then perform

Tanh activation function processing and output the image after the artistic style transfer. Specifics are as shown in Table 3.

As a result of artistic style transfer, some grid-like images often appear. This is because the images of artistic style transfer are generally low-scoring and have high-level semantic information. After the features of the image are extracted through the convolutional neural network, the features are combined to complete the artistic style transfer of the image.

4.1.3. Discriminant Network Design. The discriminant network in the CycleGAN network uses a convolutional neural network to extract the features of the entire image and then based on the extraction to distinguish the features. However, this method is easy to ignore the detailed information of the image, resulting in problems such as blurring and loss of details in the generated image. This paper uses the multiscale discriminant network structure in the construction of the discriminant network. The multiscale discriminant network structure is an improvement of the block discriminant network structure. The block discrimination network structure is to discriminate image blocks, and finally, take the average value as the final discrimination result of the discrimination network. Compared with discriminating the entire image, the block discrimination network makes the generated image closer to reality.

The multiscale discriminant network combines multiple block discriminant networks to enhance the discriminative

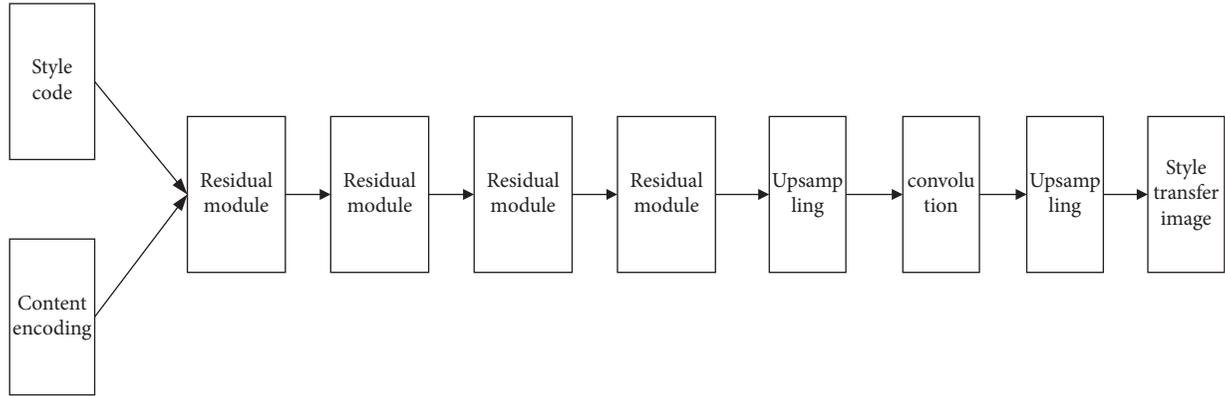


FIGURE 7: Generate network structure diagram.

TABLE 3: Generate network information table.

Operate	Kernel	Strides	Padding	Norm	Activation
Resblock1	(3,3,256)	1	1	Adaptive_Instance_Norm	ReLU
Resblock2	(3,3,256)	1	1	Adaptive_Instance_Norm	ReLU
Resblock3	(3,3,256)	1	1	Adaptive_Instance_Norm	ReLU
Resblock4	(3,3,256)	1	1	Adaptive_Instance_Norm	ReLU
up_Sample	—	—	—	—	—
Conv	(5,5,128)	1	2	Layer_Norm	ReLU
up_Sample	—	—	—	—	—
Conv	(5,5,64)	1	2	Layer_Norm	ReLU
Conv	(7,7,3)	1	3	—	Tanh

ability of the discriminant network. In this way, this article promotes the generated images to be more realistic, thereby improving the quality of the generated images.

It is judged that the network structure of networks D1 and D2 is the same. In order to accelerate the convergence speed of the network and improve the generalization performance of the model, to make the image of the artistic style transfer more lifelike, the multiscale discriminator used in this article is composed of two identical discriminators, and each discriminator model has the same network structure. Discrimination, using the discriminant results of image blocks of different sizes as the final output, thereby improving the discriminative ability of the discriminant network. The discriminant network used in this paper is shown in Figure 8.

In the final analysis, the discriminant network is a convolutional neural network composed of a convolutional layer and a fully connected layer. The main principle is to compare and analyze the features of the sample images. The specific structure diagram is shown in Figure 9.

The network structure of the discriminant network consists of 4 convolutional layers and 1 fully connected layer. The specific process is described as follows:

- (1) The input image is a three-channel image with a size of 256×256 .
- (2) The first to fourth layers all use 4×4 convolution kernels, the number of convolution kernels are 64, 128, 256, and 512, the sliding step size is 2 for both,

the filling size is 1, and the LReLU activation process is performed.

- (3) The fifth layer uses a 1×1 convolution kernel, the number of convolution kernels is 1, and the sliding step is 1, specifics are as shown in Table 4.

As shown in Table 4, in order to prevent the problem of gradient disappearance during the training process of the generative adversarial network, this paper uses the LReLU activation function in the activation function of the discriminant network except for the output layer. The last layer uses a 1×1 size convolution kernel to realize the function of full connection, convert multi-dimensional input data into one-dimensional, and output the discriminant result of the discriminant network.

4.2. Design of Loss Function. The CycleGAN total loss consists of three parts: two adversarial losses and a cycle consistent loss. Taking $x \rightarrow G(x) \rightarrow F(G(x))$ as an example, the $x \rightarrow G(x)$ adversarial loss is

$$l_{GAN}(G, D_y) = E_{y \sim p_{data}(y)} [\log D_y(y)] + E_{x \sim p_{data}(x)} [\log(1 - D(G(z^{(i)})))] \quad (1)$$

The confrontation loss of $G(x) \rightarrow F(G(x))$ is

In the training process of the network, the loss function can speed up the training of the network and improve the quality of image art style transfer.

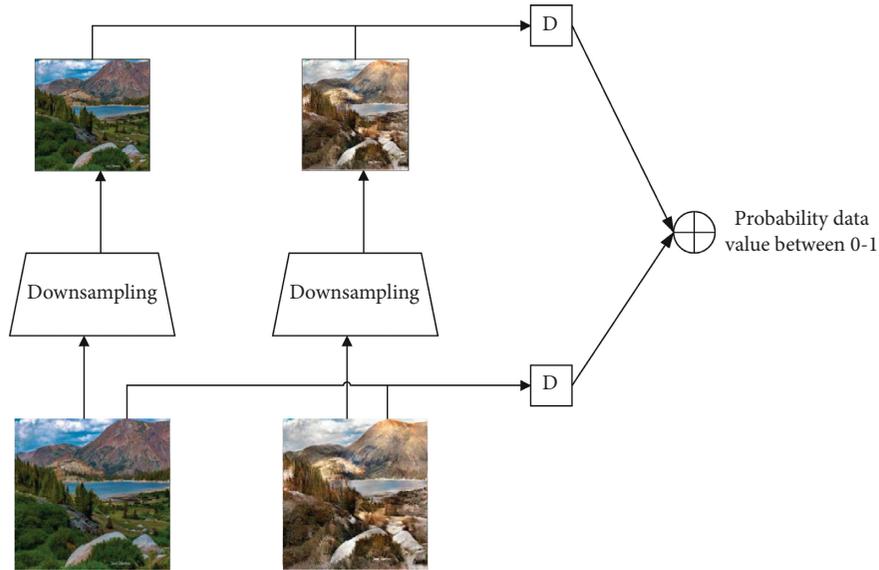


FIGURE 8: Discrimination network.

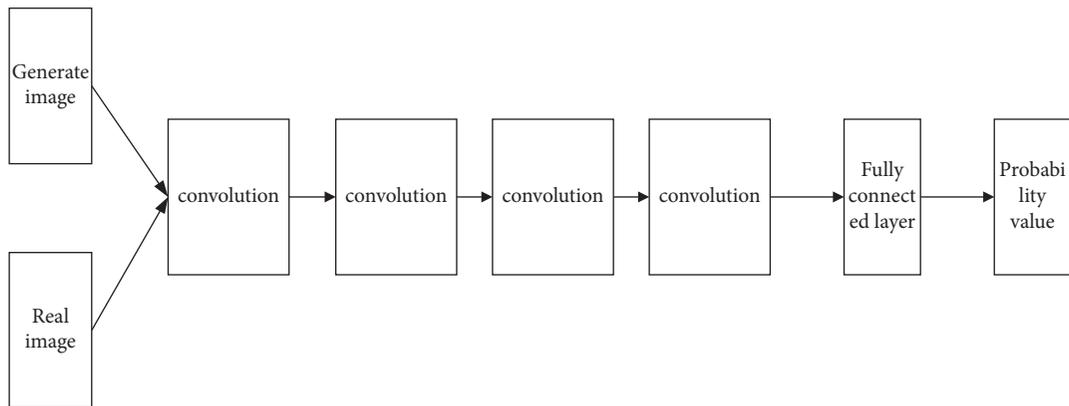


FIGURE 9: Discrimination network structure diagram.

5. Experiment and Result Analysis

5.1. Experimental Environment. The article mainly uses Python to write the source code for generating the confrontation network model. The entire deep learning framework is mainly implemented by TensorFlow. The example chosen this time is to migrate the artistic style of different images in summer and winter. The database is summer2winter_yosemite. The entire database has 4482 photos in summer and winter. The training set divides the photos into two folders trainA and trainB according to 8 : 2; the test set is divided into testA folders and testB folders. In the experiment, the batch size is set to 1, and the epoch is set to 10. Through repeated iterations, the Adam optimization parameters learning rate is set to 0.001 for optimization [24].

In the experiment established in this paper, the parameters of the image reconstruction loss function and the content encoding loss function are set, respectively. The weight of the image reconstruction loss function λ_x is set to 10.0, the weight λ_c of the content encoding loss function is

set to 1.0, and the cyclic consistency loss weight of the function λ_{cyc} is set to 10.0, and the counter loss function λ_{gan} is set to 1.0, as shown in Table 5.

5.2. Experimental Results and Analysis. In the process of running the model, you need to save the encoder, generate a network, discriminant network parameter settings, and model running parameter settings in a file every 30 times of training. When the training is interrupted due to external factors, the model has a memory function, and it can continue to run from the interrupted position until the training is completed. After all network training is completed, the artistic style is transferred through the saved model file information. Figure 10 is the winter image after the summer image migration experimental results of artistic style photos.

After testing the image art style migration of the summer2winter_yosemite dataset, it can be seen from Figure 10 that the algorithm model established in this

TABLE 4: Discrimination network information table.

Convolution number	Kernel	Strides	Padding	Activation
Conv	(4,4,64)	2	1	LReLU
Conv	(4,4,128)	2	1	LReLU
Conv	(4,4,256)	2	1	LReLU
Conv	(4,4,512)	2	1	LReLU
Fc	(1,1,1)	1	—	—

TABLE 5: Loss function weight settings.

Parameter	value
λ_x	10
λ_c	1
λ_s	1
λ_{gan}	1
λ_{cyc}	10

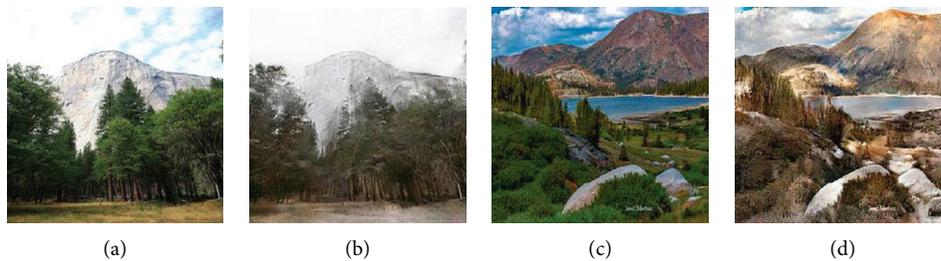


FIGURE 10: Summer to winter experiment. (a) Original image. (b) Image after artistic style transfer. (c) Original image. (d) Image after artistic style transfer.

paper realizes the conversion and fusion of summer and winter image styles, and the converted images can also match The experimental results of the summer and winter images in the dataset.

6. Conclusion

Image art transfer technology has now been widely used and developed prospects, and has relatively successful cases in many fields, but also attracts a large number of experts and scholars to study. This paper proposes an improved generative adversarial network model, which improves the traditional drawbacks of artistic style transfer due to blurred images. The method can extract the content and style characteristics of the selected data set images at the same time. Based on the predecessors, the traditional generative confrontation network model is improved. The content and style encoder of the model are used to extract the features of the image under study. The automatic encoder based on the neural network applies the normalization method to the content and style features. Fusion changes the deconvolution operation in the traditional CycleGAN network, uses image size and convolution, improves the traditional discriminant network, and uses multiscale discriminators to further improve the overall effect of artistic style fusion and migration.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] T. Sun, Y. Wang, J. Yang, and X. Hu, "Convolution neural networks with two pathways for image style recognition," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4102–4113, 2017.
- [2] P. Rajpurkar, J. Irvin, and K. Zhu, "Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning," 2017, <https://arxiv.org/abs/1711.05225>.
- [3] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4401–4410, Long Beach, CA, USA, June 2019.
- [4] C. Carlos, D. Soham, and X. T. Han, "Son of zorn's lemma: targeted style transfer using instance-aware semantic segmentation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1348–1352, New Orleans, LA, USA, March 2017.
- [5] R. Manuel, D. Alexey, and B. Thomas, "Artistic style transfer for videos," in *Proceedings of the German Conference on*

- Pattern Recognition(GCPR)*, LNCS, vol. 9796, Tübingen, Germany, October 2016.
- [6] R. Alec, M. Luke, and C. Soumith, “Unsupervised representation learning with deepconvolutional generative adversarial networks,” 2015, <https://arxiv.org/abs/1511.06434>.
- [7] C. Ledig, L. Theis, and F. Huszar, “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 105–114, Honolulu, HI, USA, July 2017.
- [8] P. Isola, J. Y. Zhu, and T. Zhou, “Image-to-image Translation with Conditional Adversarial networks,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5967–5976, Honolulu, HI, USA, July 2017.
- [9] L. A. Gatys, A. S. Ecker, and M. Bethge, “Image style transfer using convolutional neuralnetworks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2414–2423, Seattle, WA, USA, December 2016.
- [10] M. Arjovsky, S. Chintala, and L. . Bottou, “Wasserstein GAN,” <https://arxiv.org/abs/1701.07875>.
- [11] A. Madry, A. Makelov, and L. Schmidt, “Towards deep learning models resistant to adversarial attacks,” 2017, <https://arxiv.org/abs/1706.06083>.
- [12] N. Akhtar and A. Mian, “Threat of adversarial attacks on deep learning in computer vision: a survey,” *IEEE Access*, vol. 6, pp. 14410–14430, 2018.
- [13] P. Rosin and J. Collomosse, *Image and Video-Based Artistic stylisation*, Springer Science& Business Media, Berlin, Germany, 2012.
- [14] R. R. Selvaraju, M. Cogswell, and A. Das, “Grad-cam: visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, Venice, Italy, October 2017.
- [15] A. Dosovitskiy and T. Brox, “Inverting visual representations with convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4829–4837, Las Vegas, NV, USA, June 2016.
- [16] L. A. Gatys, A. S. Ecker, and M. Bethge, “A neural algorithm of artistic style,” vol. 1508, 2015 <https://arxiv.org/abs/1508.06576>, Article ID 06576.
- [17] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, <https://arxiv.org/abs/1409.1556>.
- [18] K. He, X. Zhang, and S. Ren, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [19] C. Li and M. Wand, “Combining Markov random fields and convolutional neural networks for image synthesis,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2479–2486, Las Vegas, NV, USA, June 2016.
- [20] L. Yu, W. Zhang, and J. Wang, “SeqGAN: sequence generative adversarial nets with policy gradient,” <https://arxiv.org/abs/1609.05473>.
- [21] J. Li, W. Monroe, and T. Shi, “Adversarial learning for neural dialogue generation,” <https://arxiv.org/abs/1701.06547>.
- [22] L. A. Gatys, A. S. Ecker, and B. Matthias, “A neural algorithm of artistic style,” <https://arxiv.org/abs/1508.06576>.
- [23] A. Selim, M. Elgharib, and L. Doyle, “Painting style transfer for head portraits using convolutional neural networks,” *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1–18, 2016.
- [24] A. Mahendran and A. Vedaldi, “Understanding deep image representations by inverting them,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5188–5196, Boston, MA, USA, June 2015.