Deep Learning-Based Application of Image Style Transfer

YiMi Liao and YouFu Huang

Guangzhou Panyu Polytechnic, Guangzhou, China

Correspondence should be addressed to YiMi Liao; liaoymm@qq.com

Received 24 November 2021; Revised 11 May 2022; Accepted 23 July 2022; Published 23 August 2022

Academic Editor: George Papakostas

Copyright © 2022 YiMi Liao and YouFu Huang. 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.

Generative adversarial network (GAN) is a deep learning model that is widely applied to image generation, semantic segmentation, superresolution tasks, and so on. CycleGAN is a new model architecture that is used for various applications in image translation. This paper mainly focuses on the CycleGAN algorithm model. To improve the network model’s capacity of extracting image features, the generator model uses the neural network of Unet, which consists of eight down-sampling and eight up-sampling layers, to extract image features. We use the Markov discriminator of PatchGAN since it has high-resolution and high-detail characteristics in the image style transfer. In order to improve the running efficiency, the depthwise separable convolution and standard convolution are combined in the Markov discriminator. The experimental results show that it can effectively shorten the running time. Then, we compare the image with the generative image that uses the $L_1$ loss function, the $L_2$ loss function, and the smooth $L_1$ loss function. The experimental results show that the CycleGAN neural network can effectively complete the image style transfer. The $L_1$ loss model can well retain the details of the original image. The $L_2$ loss model is clear in the distant part of the natural photo generated by Monet painting, and the color tone is more similar to the original image. The image generated by the smooth $L_1$ loss model is smoother. The $L_1$ loss model and the smooth $L_1$ loss model have some miscoloring in the natural photos generated by the Monet painting. In general, the $L_2$ loss model is more stable, and the generative image is better.

1. Introduction

Deep learning has achieved remarkable results in search technology, data mining, natural language processing, computer vision, and other related areas; it solves many complex pattern recognition problems through training to make machines imitate human activities, such as audiovisual and thinking.

In 2014, Goodfellow et al. proposed an unsupervised network learning model of generative adversarial network (GAN) [1], and the model consists of a generative network and a discriminant network. The alternate optimization between the two networks can form an adversarial process. With the improvement of the discriminant network, the data generated by the generative network are close to the distribution of the real data. Therefore, the generative network has better capabilities than other networks in data generation. This algorithm has a wide range of applications in natural language processing, image generation [2], semantic segmentation [3], image denoising, image superresolution [4], image completion, image style transfer, and so on.

In 2016, Isola et al. [5] proposed the pix2pix model as a representative work of image-to-image translation. The algorithm uses many paired images for supervised training, and it obtains a one-to-one image translation network, which can accomplish the task of image style transfer excellently. Although pix2pix can achieve realistic image conversion, the training of this model requires a large amount of paired image data, which limits its promotion and application because there are not many paired image data sets in reality.

In 2017, Zhu et al. [6] proposed an unsupervised adversarial network of CycleGAN based on GAN to break the limitation of pix2pix. The image style transfer in GAN is the conversion of one type to another type of image, which is similar to machine translation, i.e., one language to another language. The network does not need to train the paired data set and only uses the generator and the discriminator to
complete the image domain conversion. To better preserve the image content structure, it uses cycle consistency to constrain and ensure the content of the image. CycleGAN can obtain the training model by training only two types of input images, which have a wide range of applications.

In deep neural network training, with the increase of the network depth, the network model can theoretically achieve better results. But the experimental results show that the deep neural network has a degradation problem, and the training result of the 56-layer network is worse than that of the 20-layer network. The degradation problem of deep networks at least shows that deep networks are not easy to train. He et al. [7] proposed a network model residual neural network of ResNet in 2015, which can solve the problem of excessive depth leading to difficulties in training or insufficient network training.

Meili et al. [8] proposed an image style transfer method based on semantic segmentation and applied it to the transfer of Mongolian and Han clothing images. Zhang et al. [9] use global style loss and local style loss to construct total style loss. Global and local style loss is calculated by Gram matrix and Markov random field expression feature, respectively. While preserving image macro and microimage information, the image style information and structure information is improved. Zheng and Liu [10] proposed a method that can train GAN to complete style transfer using only one image, which provides a solution for training on missing image sample data sets. Li et al. [11] proposed a GAN-UNet-based ore image segmentation method, which improves the network's ability to recognize ore edges. Chen-kui et al. [12] used style transfer samples to improve the performance of pedestrian recognition. Chen et al. [13] proposed a generative network based on the fusion model of squeeze-and-excitation networks and a discriminative network based on the squeeze-and-excitation residual network are applied to image inpainting.

This paper mainly focuses on the principle of the CycleGAN in image style transfer. In the process of building the CycleGAN model, to improve the network model's ability to extract image features, the generative model uses the Unet neural network to replace the original ResNet. UNet was first published in 2015 and is a symmetrical network structure. The original intention of Unet is to solve the problems of biomedical images, and it is widely used in medical image semantic segmentation [14, 15] tasks in the early stage because of its good effect. UNet shares more network information through jump connections, and its network model is shown in Figure 1, in which the discriminator uses a Markov discriminator of PatchGAN and uses depthwise separable convolution to improve efficiency deep separable convolution is integrated into it to improve efficiency.

For this paper, the main contributions are as follows:

1. Make full use of the cyclic structure of CycleGAN and use two constraint branches in the model.
2. The generator model uses UNET neural network instead of the original RESNET neural network to train the network.
3. Construct a discriminator network model based on standard convolution and depthwise separable convolution. This method combines the advantages of standard convolution and depth-separable convolution, which can reduce network parameters and improve efficiency.
4. Compare the image with the generative image that uses the $L_1$ loss function, the $L_2$ loss function, and the smooth $L_1$ loss function.

2. Algorithm

2.1. GAN. Currently, GAN is one of the hotspots in the field of deep learning. The simplest GAN includes a generator of $G$ and a discriminator of $D$. The generator $G$ mainly generates pictures, and its input can be random noise or photos. The discriminator $D$ judges the input picture, and its input can be a binary value or a probability distribution. In the training process, the generator strives to make the generated image closer to the distribution of the picture, whereas the discriminator identifies the image. This process is equivalent to a two-person game process. The generator and the discriminator continuously improve the capacity of generating and discriminating images via training, and finally, the two networks reach a dynamic equilibrium state. In the ideal state, if the probability distribution was the output, the probability that the prediction of a given image discriminator is true approximates 0.5 (equivalent to the category of random guessing), i.e., $D(G(z)) = 0.5$. The GAN flow chart is shown in Figure 2.

In addition to CycleGAN, GAN also has several typical applications in image transfer, such as StarGAN and AttGAN. Pix2Pix solves the problem of one-to-one image conversion, whereas CycleGAN is a conversion from one domain to another. When there are many domains to be converted, we need to retrain a model for each one. In 2017, Choi et al. [16] proposed a novel and scalable method, i.e., StarGAN, which can use a single model to perform image conversion algorithms in multiple domains. It can be applied to face attribute conversion. At the same time as StarGANZ, He et al. [17] published another paper on the multidomain transfer of face attributes, proposing the AttGAN method; the method uses a unified framework for the transfer of face attributes like StarGANZ.

2.2. Principles of CycleGAN Image Style Transfer Algorithm. CycleGAN is a GAN-based network model. Unlike the general GAN network model, where the input is random noise, the input of CycleGAN is a type of picture, including two GANs. There are two generator networks ($G_{XY}$ and $G_{YX}$) and two discriminator networks ($D_X$ and $D_Y$); unlike the pix2pix model that only trains a single paired image, CycleGAN uses a generator and a discriminator to achieve mutual mapping between the two image domains, i.e., the two types of style pictures mutual conversion. Assuming that the data set to be converted is $X$ and $Y$, the generator $G_{XY}$ converts the image of type $X$ to type $Y$, the generator $G_{YX}$ converts the image of type $Y$ to type $X$, and the discriminator $D_X$ determines whether the image is of type $X$. The
discriminator $D_Y$ is to judge whether the image is of type $Y$. Figure 3 shows the cycle structure of CycleGAN. The CycleGAN loss function consists of two parts. The first is the adversarial loss in the classic GAN network. The second is the cycle consistency loss, which maps $X$ to $Y$ and then reconstructs back to $X$, i.e., the reconstruction loss.

2.2.1. Adversarial Loss. Similar to the principle of GAN, the adversarial loss of the CycleGAN makes the data generated by the generator as approximate to the real data distribution as possible. The adversarial loss consists of two steps: the first is $G_{YX}$ to realize the generation of $Y \rightarrow X$, the training tries to make the generated data $G_{YX}(Y)$ close to the real $X$, and the loss function is given as follows:

$$L(G_{YX}, D_X, X, Y) = E_{Y \sim P_{data}(Y)}[\log D_Y(Y)] + E_{X \sim P_{data}(X)}[\log (1 - D_Y)G_{YX}(X)]. \quad (1)$$

In the same way, the second part of $G_{XY}$ is to realize the generation of $X \rightarrow Y$; the training makes the generated data $G_{XY}(X)$ as close to the real $Y$ as possible, and the loss function is given as follows:

$$L(G_{XY}, D_Y, X, Y) = E_{X \sim P_{data}(X)}[\log D_X(X)] + E_{Y \sim P_{data}(Y)}[\log (1 - D_Y)G_{XY}(Y)]. \quad (2)$$

2.2.2. Cyclic Consistency Loss. Although the above-mentioned adversarial loss allows the generators $G_{XY}$ and $G_{YX}$ to learn the distribution of the data domain $Y$ and $X$, there is no guarantee that the content of the image $G_{XY}(X)$ will not change obtained from $X$. Because the data domain generated by the generator $G_{XY}(X)$ only conform to the distribution of $Y$, $X$ to $G_{XY}(X)$ contains many possible mappings. Therefore, we used the cyclic consistency loss as a constraint, and the $G_{XY}(X)$ generated by the generator can still be consistent with $Y$ in content. The representation of cyclic consistency is as follows: after $X$ passes through the generator $G_{XY}(X)$, $G_{XY}(X)$ is sent to the generator $G_{YX}$ again to obtain reconstruction at the same time, the $X^* = G_{YX}(G_{XY}(X))$ of the constraint reconstruction is as close to $X$ as possible. The principle of cyclic consistency is written as

$$L(G_{XY}, G_{YX}) = E_{Y \sim P_{data}(Y)}[\|G_{XY}(G_{YX}(Y)) - Y\|_1]. \quad (3)$$

In the same way, after $Y$ passes through the generator $G_{YX}(Y)$, $G_{YX}(Y)$ is sent to the generator $G_{XY}$ again to obtain the reconstructed $Y'$, while we constrain the reconstructed $Y' = G_{XY}(G_{YX}(Y))$ to be as close to the distribution of $Y$ as possible, and its formula is given by
L(G_{XY}, G_{YX}) = E_{X \sim P_{data}(X)}\left[\|G_{XY}(G_{YX}(X)) - X\|_{L_1}\right]. \quad (4)

The total loss of cyclic consistency is the sum of the two, and its formula is given by

\[ L_{cycle}(G_{XY}, G_{YX}) = L(G_{XY}, G_{YX}) + L(G_{YX}, G_{XY}). \quad (5) \]

2.2.3. Total Loss Function. The sum of the adversarial loss and cycle consistency loss constitutes the total loss of the CycleGAN. The formula is given as follows:

\[ L(G_{XY}, G_{YX}, D_X, D_Y) = L(G_{XY}, D_X, X, Y) + L(G_{YX}, D_Y, X, Y) + \lambda L_{cycle}(G_{XY}, G_{YX}), \quad (6) \]

where \( \lambda \) is the weight ratio of the total cycle consistency loss to the adversarial loss. During the training process, the value of \( \lambda \) can be adjusted to optimize the output of the network training results.

2.2.4. Identity Loss. The generator \( G_{XY} \) generates \( Y \)-style images, i.e., \( Y \) passes through \( G_{XY} \) and should not be changed. In this way, we can prove that \( G_{XY} \) is capable of generating \( Y \) style. Therefore, \( G_{XY}(Y) \) and \( Y \) should be as close as possible. To maintain the integrity of the image and avoid excessive transfer, Zhu et al. [6] proposed identity loss in the paper. The experimental results show that if the loss is absent, the generator may autonomously modify the color tone of the image so that the overall color changes. The identity loss function formula is given by

\[ L_{identity}(G_{XY}, G_{YX}) = E_{Y \sim P_{data}(Y)}\left[\|G_{XY}(Y) - Y\|_{L_1}\right] + E_{X \sim P_{data}(X)}\left[\|G_{YX}(X) - X\|_{L_1}\right]. \quad (7) \]

As shown in Figure 4, the pictures are the results obtained by the CycleGAN neural network iterating for 100 epochs. Figure 4(a) is the original image, and Figure 4(b) is the image generated without adding identity loss. Figure 4(c) is the image generated by adding the identity loss. We can see that the overall color of the image generated without the identity loss greatly changes, and the image generated by adding the identity loss retains the overall tone of the original image. In the Monet painting and nature photos training set, we do not need to change the overall color of the input image; therefore, the addition of the identity loss is better for image generation.

2.2.5. \( L_1 \) Loss, \( L_2 \) Loss, and Smooth \( L_1 \) Loss. To make the tone of the generated image as close to the real image as possible, CycleGAN uses the \( L_1 \) loss function to compare the generated image with the real image. Both the \( L_1 \) loss function and the \( L_2 \) loss function are used to describe the minimized error, and \( L_1 \) is the sum of the absolute difference between the output value and the input value; the gradient value is constant. Therefore, when the learning rate is constant, the loss function fluctuates in the vicinity of the stable value, and it is difficult to reach the state of convergence, whereas \( L_2 \) is the sum of the squared difference between output value and input value, when an abnormal value occurs, due to the squared difference, it will cause a greater error. Smooth \( L_1 \) combines the advantages of \( L_1 \) and \( L_2 \), and its calculation formula is given as follows:

\[ \text{smooth } L_1 = \begin{cases} \frac{1}{2} |G(i, j) - F(i, j)|^2, & |G(i, j) - F(i, j)| < 1, \\ |G(i, j) - F(i, j)| - \frac{1}{2}, & |G(i, j) - F(i, j)| \geq 1, \end{cases} \quad (8) \]

where \( G \) is the original image, and \( F \) is the image generated by the GAN. Figure 5 shows the \( L_1 \) loss function, the \( L_2 \) loss function, and the smooth \( L_1 \) loss function. We can see that when the difference between the output value and the input value is small, the gradient value of smooth \( L_1 \) loss function will still change slightly, which solves the nonsmooth problem of \( L_1 \); and when the output value and the input value differ greatly, the gradient explosion problem will not occur.

2.2.6. Depthwise Separable Convolution. Depthwise separable convolution [18] is essentially a branch network structure, which decomposes the standard convolution into depthwise convolution and point convolution. The specific operation is first to do independent depthwise convolution for each input channel and then do conventional convolution through the 1 * 1 convolution window to map the output of depthwise convolution to the new channel. Therefore, depthwise separable convolution can reduce the amount of parameters and calculation and improve efficiency, but it may also cause gradient loss.

3. Network Design

3.1. Generative Model. In building network model, the traditional CycleGAN mainly uses the deep ResNet to train the parameters of the network. The ResNet is mainly used to solve the problem of decreasing gradient caused by gradient explosion or gradient disappearance, helping to train deeper network. To improve the ability of the network model to extract image features, the generator model uses Unet neural network to replace the original ResNet. Unet network is a semantic segmentation convolutional neural network, and its typical feature is U-shaped symmetry structure, the left side is the convolutional layer, the right side is the up-sampling layer. At the same time, the feature maps obtained by each convolutional layer of the Unet network will combine the features with the corresponding up-sampling layer so that the feature maps of each layer can be effectively used in subsequent calculations, i.e., jump connection. Compared with other network structures, the Unet network structure combines the information of low-level feature maps in extracting network features. Consequently, the resulting feature maps include both the high-level semantic information of the image and the image’s detailed features,
which can improve the accuracy of the model results. Figure 6 shows the network model of the Unet generator.

The input of the generator algorithm is an image, and the output is also an image. This structure is similar to the form of semantic segmentation; therefore, it needs to be down-sampled and then up-sampled. The Unet network designed in this paper has eight layers, mainly consisting of eight down-sampling, eight up-sampling, and jump connection, to strengthen the extraction of overview features and detailed features of the image. The left side of the network structure is the compression process, i.e., the encoder, which reduces the size of the image through convolution and down-sampling. The size of the convolution kernel is 4 \times 4, and the step size is 2; therefore, once it is down-sampled, the picture is reduced to one-half of the original.

The right part is the decoding process, i.e., decoder, which uses convolution and up-sampling to obtain the characteristics of the image. The size of the convolution kernel is also 4 \times 4, and the step size of the convolution is 2, so once the next sample layer is passed, the picture is twice the original.

In both down-sampling and up-sampling, the LeakyReLU function is used as the activation function to solve the neuron “death” problem, the same filling method is used, and dropout is used to alleviate the overfitting phenomenon of the image. To make the data obey the standard normal distribution, use the instance normalization function to normalize the data.

3.2. Discriminator Model Based on Depthwise Separable Convolution. The Markov discriminator of PatchGAN is one of discriminant models; unlike the CNN classifiers that will introduce a fully connected layer at the end, PatchGAN is completely composed of convolutional layers. Different from the ordinary GAN discriminator, the ordinary GAN discriminator maps the input to a real number, i.e., the probability that the input sample is a true sample. PatchGAN maps the input to a patch (matrix) \( X \) of \( N \times N \). The value of \( X_{ij} \) represents the probability that each patch is a true sample. Patch represents a receptive field of the input image;
therefore, the value of $X_{ij}$ represents the discriminatory output of the receptive field of the input image, and finally, $X_{ij}$ is averaged as the true or false output of the image, which is the final output of the discriminator. The transfer includes two parts: content transfer and texture transfer, and the Markov discriminator can maintain the high-resolution and detail of the image to a large extent. The Markov discriminator model designed in this experiment is shown in Figure 7.

The size of the input image is $256 \times 256 \times 3$, and the convolution extracts the features of the image at the beginning. The size of the convolution kernel is $3 \times 3$, and the step size is 2. We use the same filling method as the generator, utilize dropout to alleviate the overfitting of the image, and use instance normalization to batch normalize the data in a batch. Since the information of each pixel of an image is important for image migration tasks, the instance normalization algorithm is a normalization algorithm that is more suitable for scenes with higher requirements for a single pixel. The final output is a $30 \times 30$ matrix, and the mean value of the output matrix is taken as the output of true/false. The specific network parameters of each convolutional layer of the discriminator model are shown in Table 1; four down-sampling are used: conv1, conv2, and conv3 use a $4 \times 4$ convolution kernel with a step size of 2, conv4 uses a $3 \times 3$ convolution kernel, uses the ReLU activation function, and finally outputs a $30 \times 30 \times 1$ vector. The parameters of the discriminator model are shown in Table 2.

Since CycleGAN implements mutual mapping between two image domains, and the sample set is large, in order to reduce computation, the discriminantor model in this paper replaces the standard convolution with depthwise separable convolution in conv2 and conv3. The last layer conv4 uses standard convolution. Through experiments, it can be found that the discriminator model incorporating depthwise separable convolution can effectively improve the operation efficiency.

### 4. Experimental Results and Analysis

The computer configuration used in this experiment is as follows: the processor is AMD Ryzen 7 3700X, and the operating system is Windows 10 x64. The data set is the kaggle competition platform monet2photo. The monet2photo data set includes two types of images: one is Monet painting and the other one is natural photos; there is no one-to-one correspondence between the two. The training set includes 1042 Monet paintings and 6257 nature photos, and the test set includes 91 Monet paintings and 721 nature photos.

The development tool uses PyCharm Community Edition 2019 and builds the Anaconda 3 development environment. Anaconda is an open-source Python release version, which is an open-source package and environment manager. Developers can install different versions of software packages on the same machine, and it can be switched between different environments. The main dependency packages are tensorflow2.1, Python3.6, Keras2.1, and so on. To make the model better train the data images, the image size is uniformly adjusted to $256 \times 256 \times 3$ when it is sent to the neural network, and the initial learning rate is set to 0.001. In the network training process, a phenomenon is prone to appear: the network model obtains ideal results on the training data, i.e., the loss function is small, and the prediction accuracy is high; but the
prediction accuracy on the test data is low, and the loss function is larger, which is the phenomenon of overfitting. To reduce the overfitting phenomenon of neural networks, the dropout algorithm is usually used to solve the problem. The dropout rate is set to 0.4, and the Adam stochastic gradient descent algorithm is used to optimize the network training parameters. Finally, the network model parameters are obtained, and the output model can predict the image style transfer of the newly input pictures. The parameter values are set as shown in Table 3.

Figure 8 shows the loss curve of the generator and discriminator model during the network model training process. It can be concluded that the loss of the discriminator tends to decrease with the increase of iteration during the continuous confrontation between the generator and the discriminator, and the generator loss first drops and then rises in the iterative process and tends to a stable state. This happens because the data sets trained by CycleGAN are not paired. At the beginning of the iteration, the discriminator’s ability to discriminate is weak, and the discriminator cannot distinguish the true or false of the image, but as the iteration increases, the discrimination ability improves. The loss of discriminator decreases, and it can be seen that the discriminator loss approaches 0.5. In the GAN, the loss of the generator does not represent the quality of the image generated by the generator, and the evaluation of the quality of the generated image is usually subjective. In the GAN, a reasonable evaluation of the quality of the generated image is mainly carried out from two aspects: qualitative evaluation and quantitative evaluation [19, 20]. Qualitative evaluation is mainly judged by human eyes, but the evaluation results are inconsistent due to human subjectivity. Quantitative assessments generally include IS (inception score) and FID (Fréchet Inception Distance).

We load the saved training model and input the test image for testing. Figure 9 is the effect of natural photos generated by Monet painting. Figure 9(a) is the original image, Figure 9(b) is the effect of the $L_1$ loss model, Figure 9(c) is the effect diagram of the $L_2$ loss model, and Figure 9(d) is the effect diagram of the smooth $L_1$ loss model. It can be seen that the image generated by the smooth $L_1$ loss...
model is smoother, $L_2$ loss model generates better sky effect than $L_1$ loss model and smooth $L_1$ loss, and there is no large area of miscoloring phenomenon.

Figure 10 is the rendering of natural photos generated by Monet painting. Figure 10(a) is the original image, Figure 10(b) is the effect drawing of the $L_1$ loss model, Figure 10(c) is the effect drawing of the $L_2$ loss model, and Figure 10(d) is the rendering of the smooth $L_1$ loss model. According to the renderings, there is not much difference between the effects generated by the three models, but the $L_2$ loss model generates a clearer image perspective.

The standard convolution layer can be optimized by using the depthwise separable convolution idea to reduce the number of model parameters and improve the detection efficiency of the algorithm. In the experiment in this paper, the number of discriminator parameters is reduced from 2,765,569 to 2,151,409; the experiment proves that the deeply separable convolution has obvious advantages in the training process. In order to objectively evaluate the effect of CycleGAN generated images, peak signal-to-noise ratio (PSNR) is adopted to measure the gap between the original image and the generated image. Generally speaking, the greater the PSNR value, the better the generated image effect. And using structural similarity (SSIM) to evaluate the quality of image formation, the closer the SSIM value is to 1, generated image and original image structure is more

| Parameters     | Value | Description                                           |
|----------------|-------|-------------------------------------------------------|
| $\lambda$     | 7     | The weight ratio of the total loss of cyclic consistency to the adversarial loss |
| Epochs         | 100   | Number of training iterations                        |
| $L_\gamma$     | 0.001 | Learning rate                                         |
| Beta_1         | 0.5   | Exponential decay rate                                |
| Dropout        | 0.4   | Discard rate                                          |
| Batch_size     | 200   | Training per batch                                    |

Figure 8: Loss diagram.

Figure 9: Natural photo generation Monet painting. (a) Original picture. (b) $L_1$. (c) $L_2$. (d) Smooth $L_1$. 
similar, and generate the image quality is better. PSNR and SSIM are used to evaluate the quality of image restoration, as shown in Table 4. It can be seen that the PSNR and SSIM of $L_2$ generated images were slightly higher on the whole.

### Table 4: PSNR and SSIM of the generated image.

|                | PSNR  | SSIM  |
|----------------|-------|-------|
|                | $L_1$ | $L_2$ | Smooth | $L_1$ | $L_2$ | Smooth |
| The first picture | 15.74 | 17.95 | 16.72  | 0.5872 | 0.6562 | 0.5937 |
| The second picture | 16.98 | 20.09 | 17.58  | 0.6156 | 0.6728 | 0.6349 |
| The third picture | 18.18 | 20.32 | 17.12  | 0.7488 | 0.6640 | 0.8031 |
| The fourth picture | 18.62 | 19.63 | 18.15  | 0.7535 | 0.7640 | 0.7588 |

5. Conclusion

In the field of image processing, image style transfer is an interesting research hotspot. This paper studies the CycleGAN algorithm model of image style transfer based on GAN and uses Unet neural network instead of ResNet to train the network. Compare the image generation effect under on $L_1$ Loss, $L_2$ loss and smooth $L_1$. The experimental results show that for the Monet painting and natural photo data sets, the CycleGAN neural network can effectively complete the image style transfer, and this neural network model has a certain versatility and can be applied to image style transfer in other fields. It can be seen from the visual effect that the overall effect of the image generated by the $L_2$ loss model is better. Image transfer based on deep learning brings more imagination to image style design. Future research will try to apply GAN to image restoration, color restoration, and image removal from rain and fog. Through the game style training of the generation network and the adversarial network, the accuracy of image generation is improved.

### Data Availability

The 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.

### Acknowledgments

This study received funding support from 2020 Characteristic Innovation Scientific Research Project of Guangdong Department of Education (2020KTSCX281).

### References

[1] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, “Generative adversarial Goodfellow I,” Generative adversarial, vol. 87, p. 3258, 2017.
[2] Y. Ma, X. Xu, R. Zhang, Y. Xie, and H. Chen, “Generative adversarial network and its research progress in image generation,” Journal of Frontiers of Computer Science and Technology, vol. 15, no. 10, pp. 1795–1811, 2021.
[3] R. Zhang, S. Huang, and H. Wang, Research on Road Scene Semantic Segmentation Based on Improved CycleGAN:1-9 [2021-06-10], 2014, https://kns.cnki.net/kcms/detail/11.2127.TP.20210514.1041.010.html.
[4] W. He, C. Wang, and Z. Sun, “Super resolution reconstruction of remote sensing image based on generated countermeasure network,” Information and Control, vol. 50, no. 02, pp. 195–203, 2021.
[5] P. Isola, J. Y. Zhu, T. Zhou, and A. Efros, “Image-to-Image translation with conditional adversarial networks,” in Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Honolulu, HI, USA, 21-26 July 2017.
[6] J. Y. Zhu, T. Park, and P. Isola, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2223–2232, 2017.

[7] K. He, X. Zhang, and S. Ren, *Deep Residual Learning for Image Recognition*, IEEE, Switzerland, 2016.

[8] L. I. Meili, C. Yang, and B. Shi, “Research on image style transfer technology based on semantic segmentation,” *Computer Engineering and Applications*, vol. 56, no. 24, pp. 207–213, 2020.

[9] M.-yuL. I. U. Zhang, Q. I. N. Yue-hui, and W. U. Xu-jia, “Neural style transfer method based on laplace operator to suppress artifacts,” *Computer Science*, vol. 47, no. S2, pp. 209–214, 2020.

[10] Z. Zheng and J. Liu, *PS2S-GAN: Efficient Style Transfer Using Single Style Image*, IEEE, London, 2020.

[11] H. Li, X. Wang, C. Yang, and W. Xiong, *Ore Image Segmentation Method Based on GAN–UNet [J/OL].Control Theory & Applications:1-7, 2021, Available at: https://kns.cnki.net/kcms/detail/44.1240.TP.20210330.1221.040.html.

[12] W. Chen-kui, Y. L. Chen, and X. D. Cai, “Person Re-identification method based on image style transfer,” *Journal of Beijing University of Posts and Telecommunications*, vol. 36, pp. 1–6, 2021.

[13] Y. Chen, L. Liu, J. Tao et al., “The improved image inpainting algorithm via encoder and similarity constraint,” *The Visual Computer*, vol. 37, p. 324, 2020.

[14] O. Ronneberger, P. Fischer, and T. Brox, *U-net: Convolutional Networks for Biomedical Image Segmentation*, Springer, Cham, 2015.

[15] J. Li, X. Lin, H. Che, H. Li, and X. Qian, “Pancreas segmentation with probabilistic map guided bi-directional recurrent UNet,” *Physics in Medicine and Biology*, vol. 66, no. 11, p. 115010, 2021.

[16] Y. Choi, M. Choi, and M. Kim, “StarGAN: unified generative adversarial networks for multi-domain image-to-image translation,” in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18-23 June 2018.

[17] Z. He, W. Zuo, M. Kan, and S. Shan, “AttnGAN: facial attribute editing by only changing what you want,” *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5464–5478, 2017.

[18] L. Kaiser, A. N. Gomez, and F. Chollet, “Depthwise separable convolutions for neural machine translation,” 2017, http://arxiv.org/1706.03059.

[19] H. Chen, G. Zhang, C. Gao, and Q. Zhou, “Research progress of image style transfer based on deep learning,” *Computer Engineering and Applications*, vol. 57, no. 11, pp. 37–45, 2021.

[20] F. Chen, W. U. Feng, X. Qing, W. Ying-Ming, and Y. G. En-DecUI, “A survey about image generation with generative adversarial nets,” *Chinese Journal of Computers*, vol. 44, no. 02, pp. 347–369, 2021.