Research Article

Character Creation Model and Analysis of Figure Painting Fused with Texture Synthesis Style Transfer Algorithm

Hongli Sun [iD]

XiaMen Institute of Technoligy, School of Arts&Communication, Xiamen, FuJian 361021, China

Correspondence should be addressed to Hongli Sun; sunhongli@xit.edu.cn

Texture fusion is the process of applying the style of one style image to another content image. It is an artistic creation and image editing technology. In recent years, the rapid development of deep learning has injected new power into the field of computer vision, and a large number of image style transfer algorithms based on deep learning have been proposed. At the same time, some current character conversion algorithms based on unsupervised learning also face the loss of the content and structure of the generated characters, and at the same time, they have not learned a good face deformation effect, resulting in poor image generation effects. This paper studies the relevant research background and research significance of image style transfer methods and summarizes them in the time sequence of their development; summarizes the style transfer algorithms based on deep learning, and analyzes the advantages and disadvantages of each type of algorithm. Based on the fast style transfer algorithm, this method adds a saliency detection network and designs a saliency loss function. In the training process, the difference between the saliency map of the generated image and the content image is additionally calculated, and the saliency loss is used as a part of the total loss for iterative training. Experiments show that the stylized image generated by this algorithm can better retain the salient area of the content image and has a good visual effect. Compared with the original network, the amount of parameters of the attention mechanism is very small, and there is almost no additional burden.

1. Introduction

Since the digital image coding method was proposed, images can be converted into discrete pixel data and stored in digital storage devices [1]. The advancement of digital storage has greatly promoted the development of digital image processing technology. In recent years, with the rapid development of deep learning technology, various image processing algorithms and software developed based on deep learning technology have emerged one after another, such as filters used in photographic equipment [2]. At present, image processing theory combined with deep learning technology has attracted a lot of attention in the Internet industry, and a large number of new application fields have emerged, such as mobile phone filter algorithms and video rendering. With the rise of the Internet industry such as live streaming and the large-scale use of image

acquisition equipment such as cameras, the image style transfer method has increasingly higher theoretical research and practical application value. Image style transfer is also called image style transfer, which refers to the method of transferring the style of a reference image to another or multiple original images [3]. The style here usually refers to the subjective style, such as the artistic style of oil painting or the font style of calligraphy. At present, there are no specific mathematical models and other objective indicators to measure style. The texture of an image refers to the periodic and regular information in the image [4].

The texture synthesis algorithm generally extracts and synthesizes the texture of the target image and superimposes it on the original image to achieve the purpose of style conversion. The traditional texture synthesis method is restricted by the ability of the feature extraction algorithm and often only uses the low-level features of the image, which

makes the style transfer effect unsatisfactory [5]. Classical style transfer algorithms mostly extract image features at a global level, and the quality of the generation effect of style transfer is low. In addition, the classic style transfer method is not an end-to-end structure, and it cannot be convenient and effective in practical applications [6]. After normalizing the depth features extracted by the VGG model, the co-variance matrix is obtained, and SVD decomposition is performed on it to obtain the feature code of the reference style. This way shows that the style feature space can be distinguished from the image texture. In 2019, they used wavelet transform instead of SVD decomposition to propose a new network called PhotoWCT, which transfers the style of real images at the pixel level and greatly improves the quality of image generation after style transfer [7]. On the other hand, researchers have proposed another image normalization calculation method, which uses the width and height of the image as the normalization parameters of the original image to ensure the independence between image instances while normalizing [8]. This approach can speed up network training without affecting the transfer of styles. Subsequently, the normalization calculation method was changed, and the adaptability was increased [9–11]. The reference image features extracted by the VGG model were directly integrated into the original image in the manner of normalization parameters to achieve the effect of rapid style transfer. This proves that the image features contain the image style information, and the neural network can learn this information in some way [12–14].

This paper proposes a face makeup learning network DS-DCCGAN based on the fusion of different depth and shallow layers of makeup features. The network can learn multidimensional makeup features and transfer them to the target image to further improve the quality of makeup migration. A large number of experimental results show that the method in this paper can not only achieve realistic facial makeup learning effects in terms of visual sense but also perform well on the index plane. A new double-loop uniformly constrained face makeup learning network DCCGAN is proposed. The network refers to the idea of separation of style features and content features, uses encoders and decoders to separate and merge the makeup styles of face images and image content features, and finally constrains network training through double-loop output consistency to complete the makeup migration process. Because the network separates the makeup features and content features of the image during the training process, the network only learns makeup information when the facial shadow and posture change and retains other original information of the target image. In this paper, a large number of experiments have shown that the DCCGAN network not only has better makeup migration quality in the MT data set but also has better makeup images for face makeup images with shadows, posture changes, etc., as well as for face images outside the data set.

This paper designs two image style transfer algorithms based on neural networks. First, from the perspective of improving the quality of stylized images, an image style transfer algorithm with salient area preservation is proposed. Then, from the perspective of improving the efficiency of style transfer, a lightweight image style transfer algorithm with attention mechanism is proposed.

2. Style Transfer Algorithm Based on Texture Fusion

2.1. Image Reconstruction of Character Style. In the content image reconstruction process, assuming that the original input image is x and the reconstructed image is p, P_{ij}, and F_{ij}, respectively, represent the activation value of the location of the reconstructed image feature and the ith filter j of the input image feature in layer l, and the square of the content image reconstruction [15–17]. The error loss is

$$\text{Loss}^l_{\text{cont}}(p, x) = \sum_{i=1, j=1}^{n,n} \left(F_{ij} - P_{ij}\right)^2. \tag{1}$$

The derivative function of the content loss function is

$$\frac{\partial \text{Loss}^l_{\text{cont}}}{\partial F_{ij}} = \begin{cases} \left(F_{ij} - P_{ij}\right), & F_{ij} > 0, \\ 0, & F_{ij} < 0. \end{cases} \tag{2}$$

The style feature is more abstract and complex. It depends not only on the feature information extracted from a certain layer but also on the correlation of the features extracted by the multilayer convolutional layer [18]. Gatys found that the Gram matrix can well represent this correlation. The matrix can not only measure the characteristics of its own dimensions but also measure the relationship between them. Among them, the diagonal elements represent the information of different feature maps, and the remaining elements represent the relationship between these feature maps. Therefore, this matrix is widely used in image style transfer. The Gram matrix G belongs to $R\{N_i^*N_j\}$, and G_{ij} represents the inner product of the ith feature map and the jth feature map in layer l:

$$G_{ij} = \frac{\sum_k F_{ik} F_{jk}}{2}. \tag{3}$$

Through the Gram matrix, different sizes of style image information can be obtained, but this part of the information only contains information such as texture and no global information of the image. Therefore, to generate a new image that matches a given style image, it is necessary to minimize the mean square error of the white noise image and the Gram matrix of the given style image and iteratively optimize to obtain the final texture. Assuming that a is the input style image, x is the generated image, A_l is the style representation of the lth layer of the style image, and G_l is the style representation of the lth layer of the generated image, then the style loss of the l-th layer is

$$E_l = \frac{\sum_k \left(F_{ik} F_{jk} - G_{ij}\right)^2}{2N_l M_l}. \tag{4}$$

The total style loss is

$$\text{Loss}_{\text{style}}(a, b) = \frac{\sum_{l=0}^{n} wlE}{Q_l}, \qquad (5)$$

where wl is the weighting factor of each layer's contribution to the total loss.

The style image information extracted by different convolutional layers in the network can be visualized through image reconstruction, and the visualized result is shown in Figure 1. The low-level features correspond to the pixel features of the bottom layer in the image and reflect the local content in the image. The output of the high-level network corresponds to the high-level features of the input image, reflecting the overall style of the input image.

The pretrained VGG network can extract feature information of different scales for a given style image and content image. In order to achieve the purpose of image style conversion, input a random white noise image into the VGG network to minimize the feature difference between the white noise image and the content image at the high level of the convolutional neural network and minimize the white noise image and style image in the network. The differences in the representation of the style features of each layer are finally generated through multiple iterations to generate a stylized image.

Input the style image a into the network and store the feature results of each convolutional layer. The content image p is also input to the network, and the characteristic information l_P of the high-level network is stored. Then input a random white noise image into the network, in terms of content reconstruction, in the fourth convolutional layer, calculate the loss of content image feature l_P and white noise image feature l_F. In terms of style reconstruction, at each layer of the network, the style feature representation of the style image is calculated, and the style feature representation of the white noise image is the loss of l_G. The total loss in the iteration process is

$$\text{Loss}_{\text{total}}(p, a, b) = a\,\text{Loss}_{\text{cont}} + b\,\text{Loss}^2. \qquad (6)$$

Among them, α and β are the weighting factors of content loss and style loss, respectively. Through multiple iterations, the total loss is minimized, and finally a stylized image is obtained. The convolutional neural network is successfully used for style transfer and has achieved good results [19]. It extracts the content and style characteristics of a given style image, and constructs an image with the above two characteristics through image reconstruction. The final stylized image has a visual effect Expresses a specific style. However, this type of style transfer method based on image optimization requires multiple iterations for each generation of an image, which means that each stylization requires a lot of computing resources and is not real-time.

Each iteration of the above-mentioned image optimization-based style transfer algorithm consumes a lot of time and computing resources, especially when several content images are converted into the same style, each image reconstruction will re-extract style features, resulting in computing resources Waste. The style transfer algorithm based on model optimization solves this problem well. The
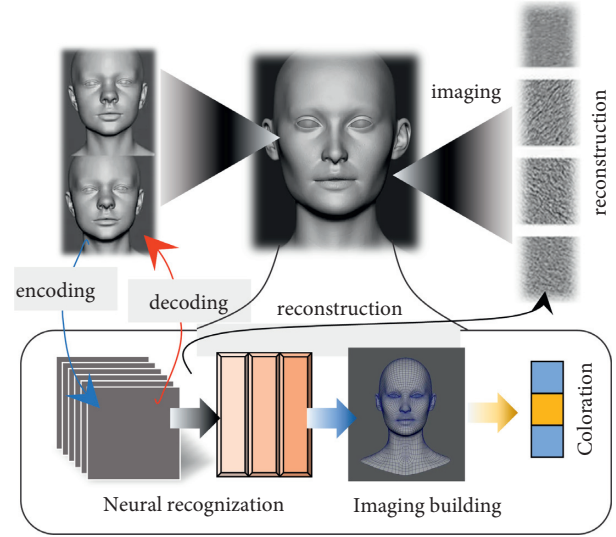

FIGURE 1: Visual reconstruction of image style.

main idea of this method is to use an image style transfer network to train different style transfer networks for different styles. The specific method is shown in Table 1. In the training phase, the content image in the data set is input into the style transfer network and the generated image is output. Then use a similar method to calculate the feature loss of the generated image and the content image in the high-level of the pre-trained VGG network, calculate the feature loss of the generated image and the style image in each layer, and then use the gradient descent strategy to adjust the network's feature loss. Parameters, so that the generated stylized image meets the expected style as much as possible. After training on massive content images, the final style transfer network model can learn the style of a given style image. In the testing phase, input the content image that needs to be converted into the trained style transfer model to quickly generate the stylized image.

In this paper, in order to solve the problem that the existing image style transfer algorithm based on convolutional neural network will cause the significant area of the content image to be distorted and lose the semantic information of the content image during the stylization process, an image style transfer algorithm with significant area preservation is proposed. . Based on the fast style transfer algorithm, this method adds a saliency detection network and designs a saliency loss. In the training process, the difference between the generated image and the saliency map of the content image is additionally calculated, so that the final generated style transfer model can retain the salient area of the content image and the semantic information of the content image while changing the style. The generated stylized image has Better visual effects.

2.2. Character Face Deconstruction.
Style transfer methods such as BeautyGAN often have unstable network training and redundant or missing facial makeup features in the makeup learning process [20]. The DCCGAN network

Data set	Data reference	Migration coefficient	Output
a1	3	4	2
a2	4	5	4
a3	6	6	2
a4	3	4	4
a5	0	2	5
a6	1	0	7
a7	2	0	6

proposed in this paper can effectively improve these problems. For example, x_1 refers to the original target domain image, and x_{21} refers to the graph x. The style of 2 is applied to the original target image, that is, the image without make-up x_1. The result of style transfer on the previous page. The network is extracted through the encoder x. Content characteristics of 1 c_1. And style characteristics s_1. Complete the task of makeup migration by exchanging the style characteristics of different images. For the image x_1. For example, after the encoder is extracted x_1. Style characteristics s_1 and content characteristics c_1. The exchange is generated by the encoder x_{12}. The two style transfer processes are combined to form a circular style transfer network.

This paper uses the characteristics of the AdaIN network and the MUNIT network to separate the image before the style transfer into image content features and style features, and then transfer the extracted style features. On the specific issue of makeup learning, the style feature parameters can be refined into the details of makeup, such as facial brightness, lipstick color, and other features. By extracting makeup features, in the case of fixed content features, it is achieved that the details, edges, and other features of the image style are preserved, and the process of makeup learning is completed. This paper uses the cyclic consistent structure to constrain the network's weakly supervised training process and uses the secondary generated image as the ground truth of the image for weakly supervised learning training.

For a single style transfer network, its structure is shown in Figure 2. The network extracts and exchanges the content feature and style feature of the image, respectively, through the content coding feature device and the style coding feature device to complete the makeup migration process. Among them, the style coding feature device is composed of a downsampling part, a global pooling layer and a fully connected layer; the content coding feature device is composed of a downsampling part and a residual module. The decoder is composed of a multilayer perceptron, a residual module, and an upsampling part. In the one-cycle generation confrontation network model, the cycle refers to the formation of a ring structure due to the symmetry of the data flow. The DCCGAN network uses the cycle consistency feature to add a new cycle on this basis to ensure the quality of the generated image of the migration result and speed up the network convergence.

2.3. Database Construction. In the training process of this network, the generator and discriminator gambling with each other, the network converges after reaching a dynamic

equilibrium [21]. The structure of the network's double loop and consistent constraint inputs the two types of images in the data set into the network at the same time to complete the style transfer process. DCCGAN implements the training of the network generator and discriminator through an optimized loss function:

$$\min \max L(a, b) = L_{\text{dynamic}} + L_{\text{equili}} + L_{\text{makeup}}. \tag{7}$$

The difference between the generated sample and the real sample can be minimized, which is equivalent to enhanced generation. The ability to generate:

$$L_{\text{dynamic}} = Ec \sim p(c_1) \frac{\log(1 - D(F(x)))}{\log(1 - D(x_2))}, \tag{8}$$

where $c \sim p(c_1)$ and $s \sim p(s_1)$ represent the input image of the generator, respectively. The content and style belong to style 1, that is, the style without makeup; it means that the image category of the discriminator belongs to style 2, that is, style with makeup. By lowering, it can maximize the generator's "spoofing" ability and at the same time enhance the discriminator's ability to discriminate images from different domains.

$$L_{\text{equili}} = L_{\text{equili}}^x + L_{\text{equili}}^y + L_{\text{equili}}^z. \tag{9}$$

The calculation method of content coding reconstruction loss, L_{reconc}, is as shown in the formula:

$$L_{\text{makeup}} = Ec \sim p(c_1) \frac{E * n_2}{\log(1 - D(x_2))}. \tag{10}$$

The above loss is back-propagated, and the style transfer network adjusts the weights according to the feedback. After multiple iterations, a style transfer model with this style is generated, as shown in Figure 3. An image style transfer algorithm for salient region preservation is proposed. Aiming at the problem that the existing style transfer algorithm will cause the salient area of the content image to be distorted during the stylization process, this paper proposes an image style transfer algorithm with salient area preservation.

Generally speaking, increasing the width and depth of the network can greatly improve the performance of the network, and the deep network is generally better than the shallow network. When simply increasing the network depth, there will be problems such as gradient explosion or gradient disappearance [22]. At this time, this problem can be solved by regular initialization, replacement of activation functions, or batch normalization. The internal residual block of the residual network uses the remaining connections, and the input image and the output image share a structure, which alleviates the problem of the disappearance of gradients caused by increasing the depth of the neural network. This kind of connection can make it easier for the network to learn the recognition ability and is easy to optimize, as shown in Table 2. This characteristic of the residual network can ensure that the input image and the output image share the structure, and maintaining a consistent image structure is beneficial for image conversion tasks.
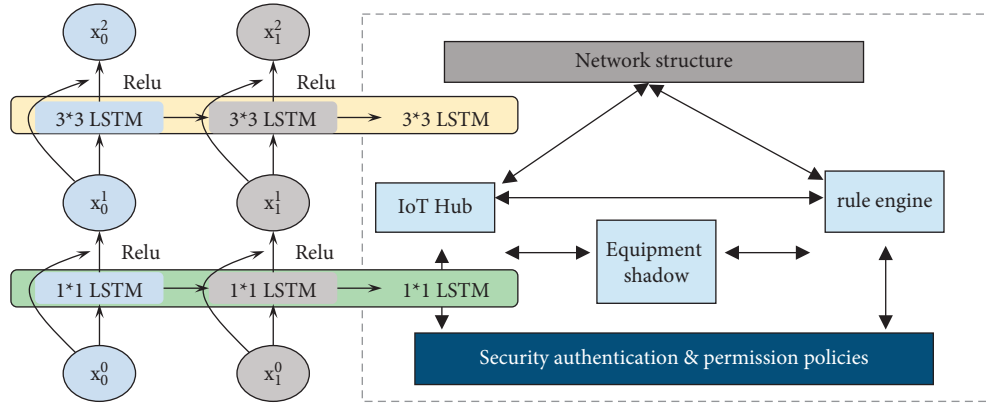
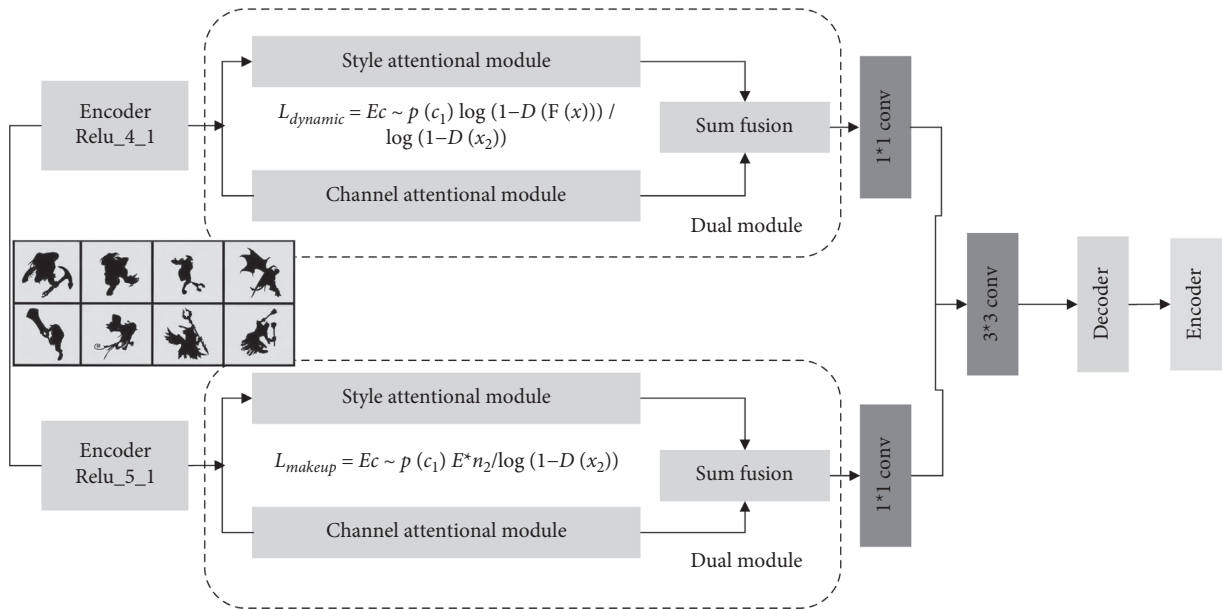FIGURE 2: Network structure of style transfer module.

FIGURE 3: Algorithm framework diagram.

TABLE 2: Comparison of conversion accuracy and saturation of residual grid.

Number of residual grids	Conversion accuracy (%)	Saturation	Consistency
1	32	65	1
2	45	74	2
3	52	78	2
4	61	87	3
5	70	89	4
6	79	96	4
7	86	97	4
8	92	97	4

The style transfer network consists of a downsampling layer, five residual blocks, and a sampling layer. The downsampling layer uses two stride convolutions with a stride of 2, which are symmetrical, and the upsampling layer uses two micro-step convolutions with a stride of 1/2. In addition to ensuring the same size of the input image and output image, the upsampling layer has two advantages: first,

a larger network can be used with the same amount of calculation, for example, a volume with a size of $3*3$ and a number of channels C. The calculation amount of the product kernel on the image of size $H * W * C$ is equal to the calculation amount of the MC channel convolution kernel in the input image size, where M is the downsampling coefficient. So this operation will greatly reduce the

computational overhead. The second is the increase in the effective receptive field. If the downsampling operation is not used, 2 receptive fields can be added for each additional $3*3$ convolution kernel. After the downsampling operation, the effective receptive field can be increased by *2M*, where *M* is the downsampling coefficient. In general, increasing the effective receptive field can improve the effect of style transfer.

Assume that the feature map size of the first layer of the VGG-16 network is $C*H*W$. *H* and *W*, respectively, represent the height and width of layer *l* features, and *C* represents the number of channels. The characteristics of the layer can be expressed as a matrix:

$$T_l \in \Re(C*H*W). \tag{11}$$

The content loss function is to calculate the mean square error between the content image and the composite image after passing through the VGG-16 network. The content loss is

$$l_{\text{cont}}(y, x*) = \frac{\left(\|T_l(y) - T_l(x*)\|\right)_2}{H*W*C}. \tag{12}$$

The style loss function calculates the square of the *F* norm of the Gram matrix difference between the style image and the composite image after passing through the VGG-16 network. The style loss is as follows:

$$T_l(y) = \|G(y) - G_l(y_s)\|^2. \tag{13}$$

For the two images in each layer of the VGG-16 network, the Gram matrix is calculated, and the square of the F norm of the matrix difference between the corresponding layers is calculated. Finally, each layer (the algorithm is ReLU1_2, ReLU2_2, ReLU3_3, ReLU4_3) adds the results to get the loss style. The Gram matrix is calculated as follows:

$$T_l(x*) = \sum_{w=1}^{n} \sum_{h=1}^{n} T_{(h,w),c}(x). \tag{14}$$

As shown in the network diagram in Figure 4, the style transfer network is trained under the guidance of the perception network and the saliency network. In the training process, the parameters of the perception network and the saliency network are fixed, and the parameters of the style transfer network are updated through the total loss function. Under the combined action of the perception network and the saliency network, multiple iterations are optimized, and the style transfer model is finally generated. This model preserves the significant area of the content image while changing the image style, and the generated stylized image has a good visual effect.

3. Results and Analysis

3.1. Saliency Evaluation of Style Transfer. It can be seen from Figure 5 that the features of the content image and the style image are extracted separately, and then the stylized image is generated through image reconstruction. The generated stylized image realizes the style conversion but is distorted,
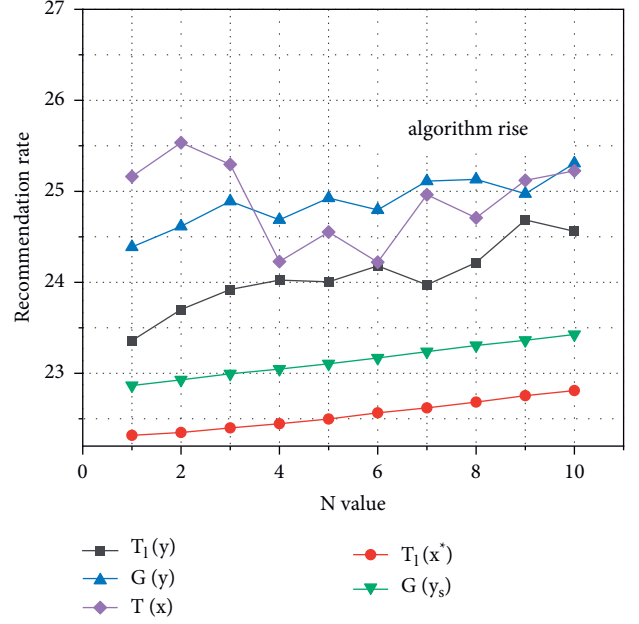

FIGURE 4: The influence of different *N* on the recommendation effect of the algorithm.
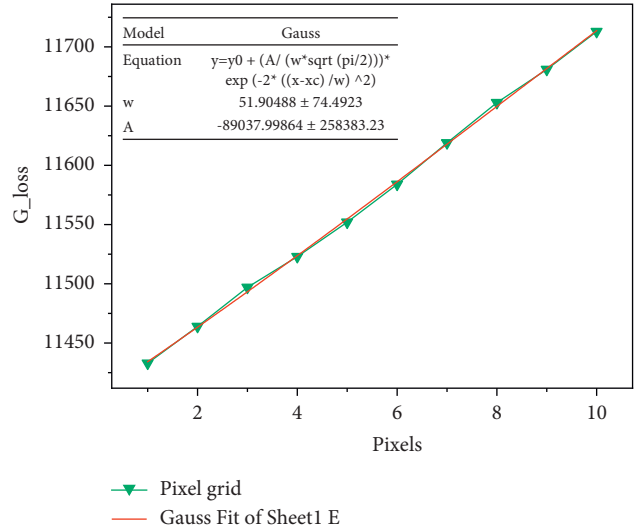

FIGURE 5: Pixel grid change curve.

and the semantic information of the content image is lost. The perceptual loss function is proposed, which retains the structure of the content image to a certain extent, but it retains all the features in the content image, so that the stylized image has no sense of hierarchy. The style transfer of any style is realized, and the generated stylized image has a relatively obvious "grid" phenomenon, resulting in unsatisfactory visual effects. The proposed image style transfer algorithm with salient area preservation can better preserve salient areas of content images in stylized images, and the salient areas are obviously different from the background and have better visual effects. From the comparison of the

saliency map of the stylized image, it can be seen that the saliency map of the stylized image generated in this chapter is the most consistent with the saliency map of the content image, which means that the stylized image generated by the algorithm in this chapter can well retain the content image while changing the style. The prominent area of the image enhances the visual effect of the style image. However, as the number of network layers further increases to hundreds of layers, the phenomenon of network degradation will occur, that is, the accuracy of the training set is saturated or decreased. In response to this problem, the literature proposed a residual network.

Through three sets of comparative experiments, it can be seen that under the same conditions, DS-DCCGAN has a better style transfer effect than DCCGAN. The DS-DCCGAN method can learn the makeup of the reference image on the basis of maintaining the facial features of the original image to be transferred, regardless of whether there is a change in facial pose or facial shadow. The calculation of the evaluation plane also uses the test set of the MT data set as input. After the models are trained using the same training set, the coordinates of the style & structure plane are calculated by Gram loss and SSIM. The comparison calculation results with the previous comparison methods are shown in Figure 6.

In the plane of he figure, the method DS-DCCGAN in this chapter is the red point, and the contrast method is the black point. The more the value goes to the upper right, the better the makeup learning effect of the image. In order to control the variables, each method in the figure uses the MT data set and the same calculation method to calculate the loss. From the ordinate point of view, the Gram loss of the method in this chapter has a better effect than DCCGAN, compared with BeautyGAN and AdaIN, the structure is closer to the original image style. This result can be understood as the method in this chapter retains more facial features of the target image under the condition of similar style transfer effects.

On the specific issue of makeup learning, the style feature parameters can be refined into the details of makeup, such as facial brightness, lipstick color, and other features. By extracting makeup features, in the case of fixed content features, it is achieved that the details, edges, and other features of the image style are preserved, and the process of makeup learning is completed. This paper uses the cyclic consistent structure to constrain the network's weakly supervised training process and uses the secondary generated image as the ground truth of the image for weakly supervised learning training.

In the evaluation index, two methods are used to judge the quality of the repaired image. One is the subjective method, that is, looking for the evaluator to give the damaged picture, the repaired picture of a certain method, and the original picture without knowing what kind of repair method the picture was obtained, and let them score the repair result (point value is 0–10, the better the repair effect, the higher the score), and then the average score of each method is counted. One is an objective way, that is, a certain algorithm index is used to characterize the similarity between the restored picture and the original picture. In this paper, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used to evaluate the quality of the restored picture. Regarding the subjective scoring method, we found 4 people unrelated to this research to score various repair methods, and each person randomly assigned 250 pairs of data in the verification set (including damaged pictures, pictures after a certain method of repair, and original pictures) to be evaluated the result, as shown in Figure 7. This method is undoubtedly the most accurate, but considering that the workload is too large and slow, this article selects PSNR and SSIM, two widely used objective evaluation standards for auxiliary evaluation.

3.2. Analysis of Experimental Results.

As we all know, the setting of hyperparameters is particularly important for deep neural networks, especially for generating adversarial networks. Therefore, in order to make the model learn as much as possible, we have done a lot of debugging experiments on the GNST-Net hyperparameters and finally selected the best hyperparameters, as shown in Figure 8:

(1) The initial training of the network does not use the discriminator, and the MSE loss without weight is used for 10 epoch training. The reason for this is to give the discriminator a better direction to "game" with the generator;

(2) The training of the generator and the discriminator is interleaved. The strategy is that when the generator is trained, if the accuracy of the batch of generated data on the global discriminator (that is, the global discriminator judges the generator generated samples as negative samples, the proportion of the total generated samples) is 30% The following is the discriminator training (including global discriminator and local discriminator), otherwise it is considered that the generator cannot hide the discriminator at this stage, and further training is required. When training the discriminator, if the accuracy of this round of the global discriminator can reach more than 75%, it is considered that the discriminator is well trained at this stage and can assist the generator training;

(3) The weight of the adversarial loss of the discriminator in the total loss when the generator is trained. The loss of the global discriminator is selected according to its size. In this paper, if the loss of the global discriminator is greater than 0.5, the weight coefficient is given as 0.25 in the Basenet_1 and Basenet_2, and if the loss is less than 0.5 in the sptnet_1 and sptnet_2, the weight coefficient is given as 0.5. For the local discriminator, the same idea as the global discriminator is adopted, but its loss is its own output.

The current face restoration algorithm has been developed rapidly, and the face restoration technology has been widely used in photo restoration and cultural relic restoration. Based on the full research and analysis of the current
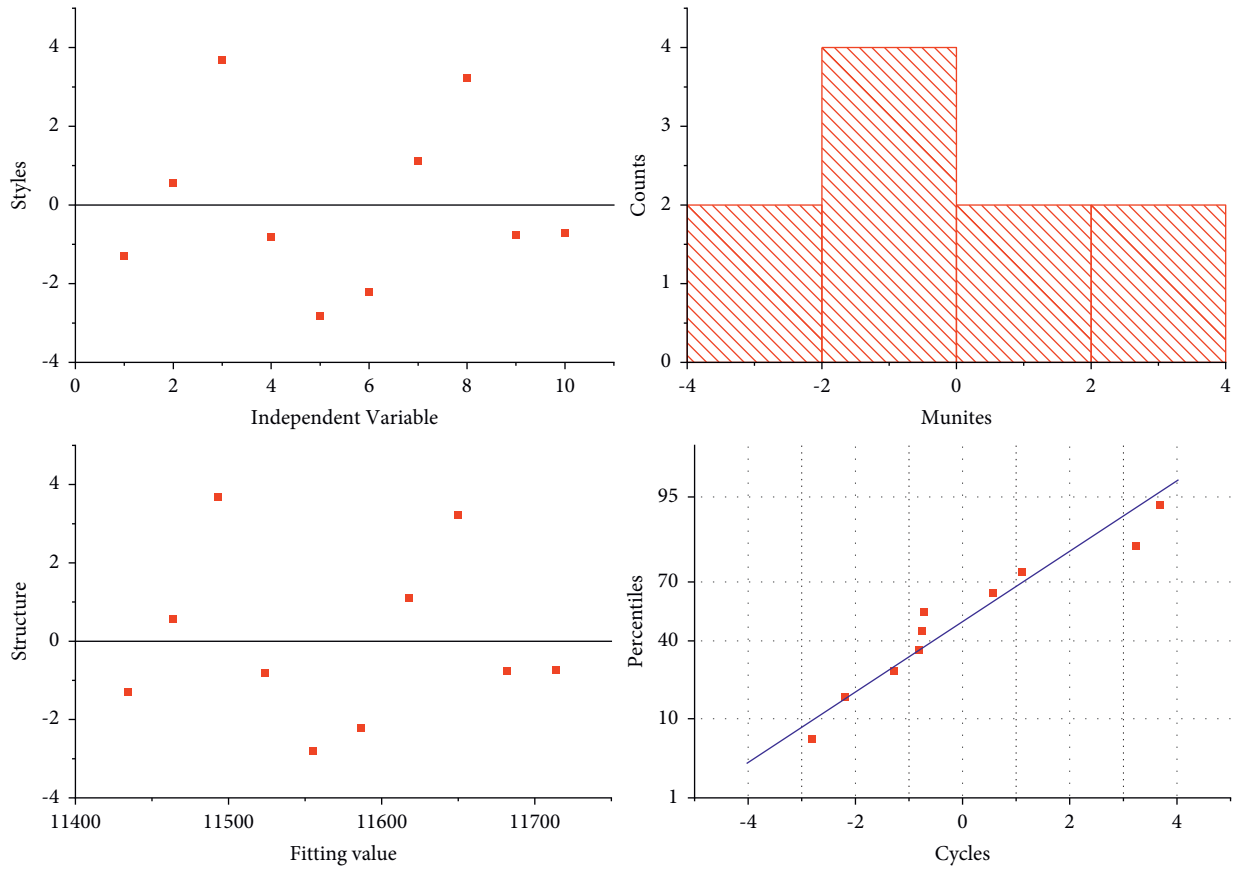
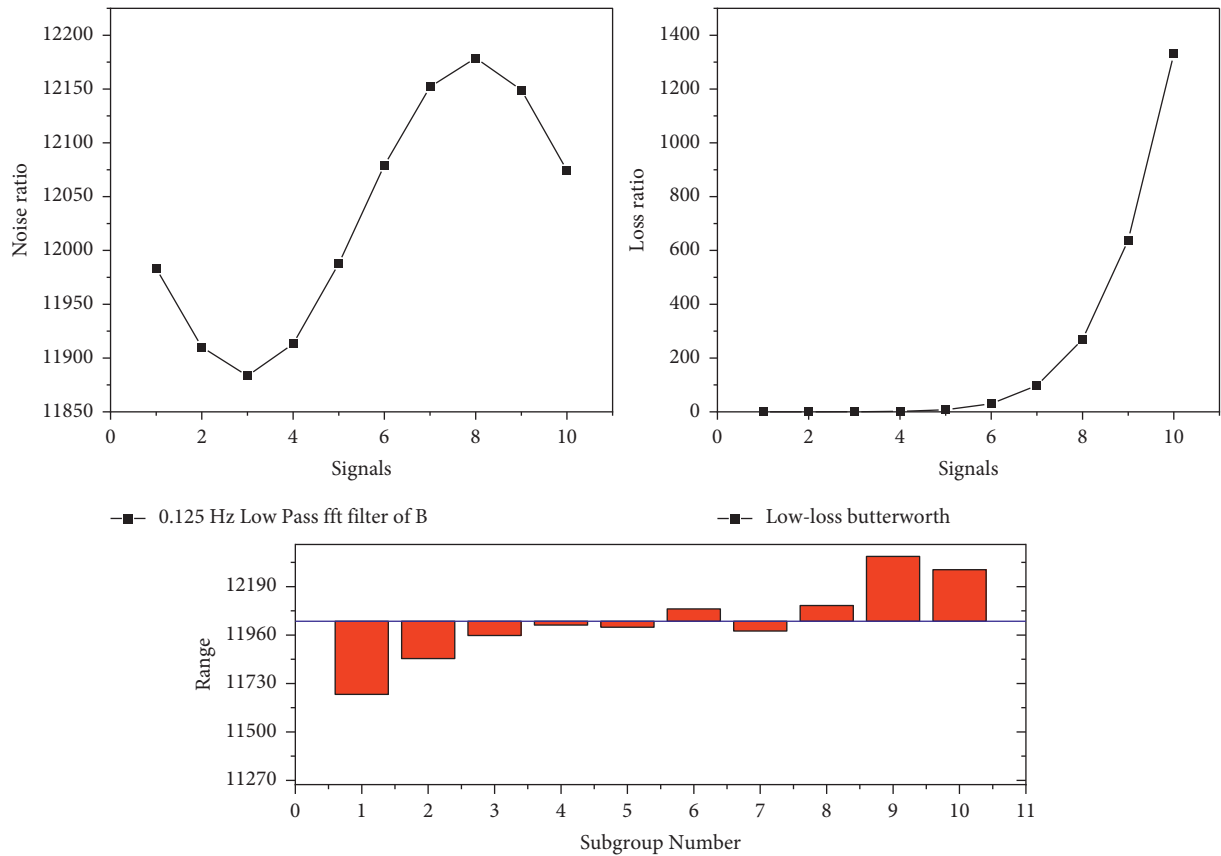Figure 6: Style and structure plan evaluation.

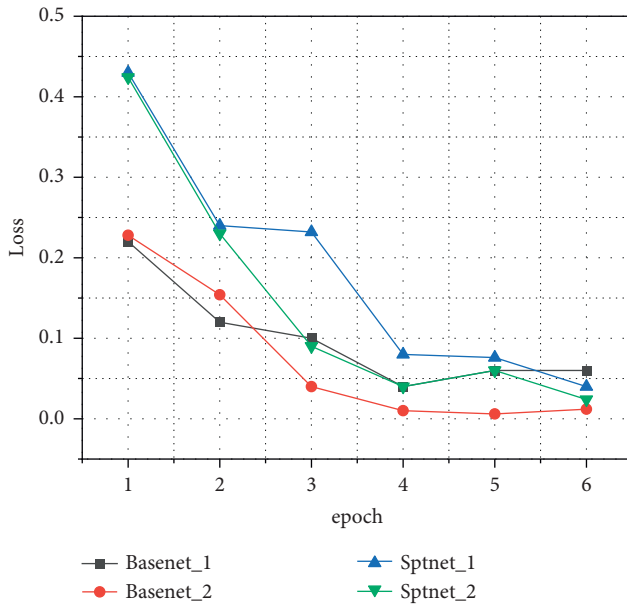Figure 7: Peak signal-to-noise ratio comparison.

FIGURE 8: Training result evaluation result.

mainstream face restoration algorithms, this paper proposes a face restoration technology based on generative confrontation network and neural style transfer. The lightweight model MobileNet-V1 is selected as the model framework of this design. Real-time performance and accuracy can be better balanced. Adding variational sampling to the generator brings more relevant information to the model, making the repaired image more detailed. Adding SptCat between encoding and decoding allows the model to reduce the cost of searching for similar blocks to a certain extent, making training faster and more stable. Adding the style loss of neural style transfer to the training makes the picture after face restoration have integrity and continuity, which improves the accuracy of face restoration to a certain extent.

4. Conclusion

Texture fusion is the process of applying the style of one style image to another content image. It is an artistic creation and image editing technology. In recent years, the rapid development of deep learning has injected new power into the field of computer vision, and a large number of image style transfer algorithms based on deep learning have been proposed. From the perspective of reducing model parameters and improving operating efficiency, this paper proposes a lightweight image style transfer algorithm that introduces an attention mechanism. Use a lightweight convolutional neural network as the main body network for style transfer and introduce a lightweight attention mechanism into the main body network. Compared with the original network, the amount of parameters of the attention mechanism is very small, and there is almost no additional burden. In this case, the performance of the lightweight network can be improved. In addition, the idea of residual network is used to add remaining connections to the lightweight network to enhance the recognition ability of the

network. Experiments show that the lightweight style transfer algorithm proposed in this chapter guarantees the quality of stylized images. The network model has fewer parameters and smaller size and is easy to apply on mobile devices. In the future, the prominent area of the image will enhance the visual effect of the style image.

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 there are conflicts of interest.

Acknowledgments

References

[1] X. Guan, L. Likang, L. Honglin et al., "Automatic embroidery texture synthesis for garment design and online display," *The Visual Computer*, vol. 37, no. 9, pp. 2553–2565, 2021.

[2] S. K. May, I. G. Johnston, P. S. C. Taçon, I. Domingo Sanz, and J. Goldhahn, "Early Australian anthropomorphs: jabiluka's dynamic figure rock paintings," *Cambridge Archaeological Journal*, vol. 28, no. 1, pp. 67–83, 2018.

[3] T. Borea, "Re-introduzione all'artista pier francesco cavazza (1677-1733)," *Tab*, vol. 4, no. 1, pp. 1–7, 2019.

[4] M. Faietti, "Davanti Al botticelli: tre ricerche sul dettaglio significante," *Figure*, vol. 4, pp. 106–129, 2019.

[5] U. K. Frederick, "Figure on a sandstone ground: considering brett whiteley's rock art," *Australian and New Zealand Journal of Audiology*, vol. 19, no. 2, pp. 168–188, 2019.

[6] D. Banerjee, "Waiting figure in a landscape: the painting of carole gibbons," *Visual Culture in Britain*, vol. 21, no. 1, pp. 114–132, 2020.

[7] I. Stancari, "Il Primo Album Fotografico Di Raffaele Belluzzi. Prime Indagini: concorsi e Pubbliche Collezioni," *Figure*, vol. 4, pp. 75–90, 2019.

[8] N. N. Prieto, "Joaquín inza: accademico d'onore presso l'Accademia clementina," *Figure*, vol. 4, pp. 40–48, 2019.

[9] V. E. Tebala, "Un Caso Studio Di Arte Pubblica: flavio Favelli e Il Murale Luigi Marulla a Cosenza," *Figure*, vol. 4, pp. 99–105, 2019.

[10] A. A Daniel and A. Dzikunu-Bansah, "Evaluation of figure drawing skills of art students in tertiary institutions. A case study of bolgatanga polytechnic," *Journal of Education and Practice*, vol. 10, no. 8, pp. 66–75, 2019.

[11] P. Isola, Z. Jun-Yan, Z. Tinghui, and E. Alexei, "Image-to-Image translation with conditional adversarial networks," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976, Honolulu, HI, USA, July 2017.

[12] J.-Y. Zhu, P. Taesung, I. Phillip, and A. E. Alexei, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, Venice, Italy, October 2017.

[13] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[15] L. Gatys, A. Ecker, and M. Bethge, "A neural algorithm of artistic style," *Journal of Vision*, vol. 16, no. 12, 2016.

[16] L. A. Gatys, S. E. Alexander, and B. Matthias, "Texture synthesis using convolutional neural networks," in *Proceedings of the 28th International Conference on Neural Information Processing Systems*, vol. 28, pp. 262–270, December 2015.

[17] B. Epshtein, O. Eyal, and W. Yonatan, "Detecting text in natural scenes with stroke width transform," in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2963–2970, San Francisco, CA, USA, June 2010.

[18] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 341–346, Los Angeles, CA, USA, 2001.

[19] A. Hertzmann, E. J. Charles, O. Nuria, C. Brian, and S. David, "Image analogies," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 327–340, Los Angeles, CA, USA, June 2001.

[20] A. Jalandoni, I. Domingo, and P. S. C. Taçon, "Testing the value of low-cost structure-from-motion (SfM) photogrammetry for metric and visual analysis of rock art," *Journal of Archaeological Science: Report*, vol. 17, pp. 605–616, 2018.

[21] P. S. C. Taçon, K. M. Sally, L. L. Ronald et al., "Maliwawa figures-a previously undescribed arnhem L and rock art style," *Australian Archaeology*, vol. 86, no. 3, pp. 208–225, 2020.

[22] J. A. Hayward, I. G. Johnston, S. K. May, and P. S. C. Taçon, "Memorialization and the stencilled rock art of mirarr country, northern Australia," *Cambridge Archaeological Journal*, vol. 28, no. 3, pp. 361–378, 2018.