Research on Fast Image Style Transformation Based on Residual Network

Nan Xue, Jin Duan, Bing Zhang, Xiaoyu Wang, Lintao Yu
Changchun University of Science and Technology, Changchun 130000, China
568421477@qq.com

Abstract. Image style migration technology refers to the conversion of an image into a similar image to a famous painting style through learning (using a convolutional neural network). The A Neural Algorithm of Artistic Style (NAAS) proposed by Gatys uses the VGG network to design a loss network, and the style migration image is obtained through repeated iterations, and the calculation speed is slow. Li Feifei introduced the residual network on the basis of Gatys, and accelerated the calculation by using the fast connection feature of the residual element. This method has a great improvement in the generation speed but still has room for improvement. This paper proposes improvements for the following two aspects: firstly, the classical residual element structure is adjusted, and the standard convolution is converted into point convolution and deep convolution, which reduces the calculation amount while ensuring the convolution effect; secondly, the loss network is simplified. The fourth and fifth layers in the model are highly consistent in structure, and the style restoration of the two layers is basically the same as the content reconstruction effect. Therefore, the fifth layer is deleted, the redundant parameters are removed, and the parameter amount is reduced. At the same time, the effect of style restoration and content reconstruction is guaranteed.

1. Introduction
When admiring the masterpieces of famous Chinese and foreign painters, I often hope that I can also create a painting with similar styles. The emergence of image style migration technology has helped people realize this desire. People can convert any photo into any style of painting, whether it is Impressionism, Fauvism, or ink painting, Chinese painting, can be achieved through style transfer technology. The schematic diagram of image style migration is shown in Figure 1:

![Fig.1 Schematic diagram of style transform](image)
(a) original image (b) style image (c) style migration image

Gatys took the lead in proposing the traditional method of image migrating A Neural Algorithm of Artistic Style (NAAS). Based on the famous target recognition network VGG, a loss network was designed, the loss function was calculated, and the gradient migration method was used to calculate
the style migration image[1]. Since each generation of an image requires multiple iterations, it brings a problem of slow running speed, especially when encountering many-to-many style migration applications, which greatly consumes running memory and may even cause process crash [2].

Li Feifei proposed the Fast Neural Style Transfer (FNST) method, which uses the residual network to design an Image Transform Net [3] and applies it to the style migration of images. This method generates a style image in advance in Image Transform Net and submits it to the loss network (consistent with the network used by Gateys) to finally generate a style migration image. In this method, the style image only needs to be calculated in Image Transform Net, unlike the method of Gatys, which requires hundreds of rounds of calculation, which greatly improves the operation speed .

Image Transform Net in FNST is designed based on the residual network. In this paper, the residual element structure is optimized, and the combination of point convolution and depth convolution is used instead of the classical convolution operation to reduce the parameter content while ensuring the performance of the network structure. At the same time, the output part of the loss network (ie VGG network) is simplified, the convolutional layer with similar functions is removed, the redundant parameters are removed, and the size loss function function is set to further reduce the parameter content of the network structure [4]. In the end, the parameters are greatly reduced while implementing the style migration.

2. Image style migration calculation overview

There are two main types of image style migration techniques. One is the NAAS first proposed by Gates. The method is based on the VGG network for loss network design. The other is the fast image style migration technology (FNST) proposed by Li Feifei et al. Based on the results of the Getys, this method adds an Image Transform Net to the front end of the loss network (the network is designed based on the residual network). The relationship between the two is shown in Figure 2:

![Fig.2 Style transform technology diagram](image)

2.1 Classic image style migration technology

The NAAS proposed by Gatys is based on the VGG network, using gradient descent, after several iterations to obtain the image after the conversion style.

Due to the excellent image feature extraction ability of the convolutional neural network, the loss network in the figure is designed based on the VGG network [5]. The initial style migration image is actually a white noise image, then the loss network extracts the original image and style image, calculate the feature map of the style transform image respectively. The difference between the original image feature and the migrated image feature is obtained, the content loss is obtained, and the style image feature and the migration image are obtained. The characteristics are poor, the style loss is obtained, and the two are combined to get the total loss. In the loss network, the gradient of the feature map is calculated by deriving the total loss function, and the iterative iteration finally results in a style migration image.
The VGG network was first proposed by Simonyan et al. and is widely used in the field of object recognition [8]. In the image style migration, the operation mode is opposite to that of VGG, that is, the input layer acquires the style feature of the image, and the output corresponds to the corresponding image.

2.2 Fast image style migration
As can be seen from Figure 2, the main improvement of the traditional method of FNST is the loss of Image Transform Net added by the network front end. The input image is generated by Image Transform Net to generate the migrated image, and the migrated image enters the loss network to calculate the content loss, the style loss and the total loss respectively. After the training is completed, the migrated image is similar in content to the input image and is similar in style to the specified masterpiece style. At this point it should be noted that since the migration image only needs to be calculated in Image Transform Net (rather than iterating as in the traditional method), the computational speed is significantly improved [6].

FNST is designed based on the residual network, and the residual element structure is shown in Figure 3:

Due to the existence of a shortcut, the residual element can implement association between any layers, so that in the process of forward propagation, the mapping of the residual network is actually a continuous operation. Compared with the traditional convolutional neural network, the residual network is undoubtedly better, and the computational complexity of the residual network is significantly reduced. This laid the foundation for a quick style migration.

3. Improvement of image style migration technology
NAAS needs to calculate the initial white noise image repeatedly to get the final style migration image, which directly leads to the migration process taking too long. The VGG network used in this method is known for its numerous parameters, especially the highly consistent blocks contain a large number of repeated parameters. In response to this situation, specific improvements are proposed in Section 3.1.

FNST introduces Image Transform Net based on residual network design, so that the migrated image can be optimized into the loss network only after calculation, and finally the style migration image is obtained. The residual element structure in this method is still not refined enough, and still contains more parameters. In response to this situation, specific improvements have been proposed in Section 3.3.

3.1 Simplification and improvement of loss network
The loss network is divided into 5 modules, including 16 convolutional layers and 5 pooling layers, combining the style loss and content loss to reconstruct the content and style of the image. Taking VGG-19 as an example, the parameter content of the three fully connected layers accounts for about 70% of all parameters [7], and in the process of style migration, the main part of the network is the
convolutional layer in the middle part of the network. Discarded the three fully connected layers. While reducing the parameter content, the calculation time is effectively reduced. The VGG network (ie, the loss network) accepts the input image for reconstruction of style and content. The process is shown in Figure 4:

In FIG. 4, a refers to a content reconstruction and style reconstruction effect diagram after Block_1 (ie, the first layer) of the network model. b, c, d, e, etc. and so on. The effect of the style and content reconstruction of each stage of a, b, c, d, and e is shown in Fig. 5.

It can be seen from Fig. 5 that the style reconstruction of the first layer only obtains the color information of the style image, and the feature information such as the rest texture is not reconstructed; and the content reconstruction of the first layer has completed the restoration of the content. As the level deepens, the migration image becomes closer to the famous painting in style, and the content deviates from the original image. In fact, the fourth and fifth layers (ie, the two layers d and e) lose some of the detailed features and only retain the high-level features on the content, while the two layers basically restore the style image in the style reconstruction. And as can be seen from Figure 8, the styles restored by the fourth and fifth layers have been highly consistent [8]. Considering that the fourth and fifth layers are similar in structure, in order to improve the operation speed, the loss network (ie, the VGG network) is simplified, and the fifth layer is completely discarded, leaving only the first four layers. The simplified VGG network is named Simpled VGG, referred to as SVGG. Due to the deletion of the fifth layer, the model after the simplification is insufficient in the restoration of the style compared to the original model.

3.2 Loss network simplification before and after style migration experiment comparison
On the same computer, the simplified network model before and after the simplification is used (the parameters of the loss of control style $\beta$ are consistently 0.5), and the image shown in Figure 6 (image size is 900 600) is style-shifted, and each time is time-consuming, record it.

By comparing the results of the simplified network model before and after the simplification, it can be seen that the simplified model SVGG has a higher speed and accuracy in the restoration of styles compared to the original model. The original model results are more consistent with the target style image, while the simplified model results are more detailed and richer in features.
Figures 6(c) and (d) have a good reduction of the style of the famous painting, and the difference in the degree of style reduction between the two is not obvious. Moreover, using the time function in Python to record the time spent in the two sets of experiments, the SVGG model spent 97 seconds on style migration, and the original model took 112 seconds.

It can be seen that the SVGG model takes less time to obtain the style migration effect similar to the initial model.

In order to verify the effectiveness of the improvement, five sets of comparative experiments were performed, and the time required to record each was as shown in Table 1.

| Tab.1 Comparison table between SVGG and original model |
|-----------------------------------------------|
| Number of experiments | SVGG | Original model |
|------------------------|------|----------------|
|                       | 1    | 2              |
|                       | 97   | 112            |
|                       | 2    | 91             |
|                       | 133  | 105            |
|                       | 3    | 87             |
|                       | 154  | 104            |
|                       | 4    | 102            |
|                       | 116  | 116            |

It can be seen that after artificially increasing the proportion of style loss, SVGG can achieve a similar degree of style reduction as the original model, while requiring less time, generally reducing by 13.6%.

### 3.3 Image Transform Net optimization of residual elements

Image Transform Net in FNST is designed based on the residual element, which is shown in Figure 7:

![Residual element structure diagram in Image Transform Net](image)

After inputting the 64-dimensional data into the 64-channel convolutional layer of the convolution kernel size, the relu excitation function is used to enter the next convolutional layer, and the layer structure is consistent with the upper layer. The number of parameters in the residual element is: 

$$3 \times 3 \times 64 \times 64 \times 2 = 73728$$

Although FNST has greatly improved the calculation speed compared with NAAS, the structure of the classical residual element is still not refined enough, and there is room for further improvement.

A new type of bottleneck design proposed by He Kaiming et al. further reduces the parameter content in the network model, thus laying a foundation for deepening the network [9]. In view of this, Image Transform Net, which is applied to the style migration by using the new residual element structure, reduces the parameters while ensuring the efficiency of image conversion. Its structure is shown in Figure 8:
In Figure 8, the convolutional layer with the size of the convolution kernel and the number of channels is 32. The function is to reduce the dimension, that is, to reduce the 64-dimensional data to 32-dimensional, and enter the convolution layer with the convolution kernel size and the channel number of 32. Enter the convolutional layer with a convolution kernel size of 64 and a number of channels to recover by convolution. The number of parameters in this structure is: 

\[ 1 \times 1 \times 32 \times 64 + 3 \times 3 \times 32 \times 32 + 1 \times 1 \times 64 \times 32 = 13312 \] 

Obviously, the parameter content in the optimized residual element is only 18% of the parameter amount (73728) in the original residual element.

3.4 Comparative experiment before and after residual element optimization

Apply the bottleneck design residual element structure to Image Transform Net, and record the method as BFNST. In order to verify that BFNST has faster calculation speed than FNST, the loss network is kept unchanged in the experiment. The front-end Image Transform Net uses the bottleneck design residual element and the traditional residual element structure respectively. The same image (image size 1280 900) was styled on the same computer. The experimental results are shown in Figure 9, and the time was recorded.

Comparing Fig. 9 (b) and (c), the degree of style reduction achieved by the two methods is basically the same. In terms of computing speed, BDFNST completed the style migration of the image in only 1.213 seconds, while FNST took 1.374 seconds. In order to verify the validity of BDFNST, five sets of comparative experiments were performed and the respective times were recorded as shown in Table 2:

| Number of experiments | 1   | 2   | 3   | 4   | 5   |
|-----------------------|-----|-----|-----|-----|-----|
| FNST                  | 1.374 | 1.229 | 1.346 | 1.744 | 1.439 |
| Unit: second          |     |     |     |     |     |

![Input image](image1.jpg) ![BDFNST result chart](image2.jpg) ![FNST result chart](image3.jpg)
As can be seen from Table 2, the average time required for BDFNST is 0.15 seconds less than FNST, and the average time consumption is reduced by 10.7%. It can be seen that BDFNST has faster calculation speed while achieving the same style migration.

4. Conclusion

Aiming at the traditional image style migration method, the SVGG model is proposed to simplify the loss network. Through research, it is found that the fourth and fifth layers of the original loss network are highly consistent in structure and parameters, and the functions achieved are consistent. Comparative experiments show that SVGG has a time-consuming reduction of 13.6% while achieving similar style reduction. Also, SVGG can achieve a higher degree of style reduction by manually adjusting the style loss factor.

The newly proposed BFNST method, on the basis of FNST, proposes an improvement on the residual element structure in Image Transform Net, and draws on the residual element structure of bottleneck design type to reduce the parameters in the network while ensuring the degree of style reduction. The content is powerful, which increases the calculation speed, and the time required for style migration is reduced by 10.7%.

It can be seen that the BFNST method after optimizing the residual element structure and simplifying the loss network model has a more efficient calculation speed while completing the image style migration.

References
[1] Gatys L A, Ecker A S, Bethge M. A Neural Algorithm of Artistic Style[J/OL]. arXiv:1508.06576. (2015-09-02)[2019-2-23]. https://arxiv.org/abs/1508.06576.
[2] Du X Y. Research and application on artificial intelligence of Chinese calligraphy[D]. Hangzhou: Zhejiang University, 2018.
[3] Justin Johnson, Alexandre Alahi, Li Fei-Fei. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[J]. 2016:694-711.
[4] Howard A G, Zhu M, Chen B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. 2017.
[5] Guo T H. Research on image color style migration technology for mobile applications[D]. Beijing: North China University of Technology, 2016.
[6] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J/OL]. arXiv:1409.1556. (2015-4-10)[2019-2-23]. https://arxiv.org/abs/1409.1556.
[7] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J/OL]. arXiv:1512.03385. (2015-12-10)[2019-2-23]. https://arxiv.org/abs/1512.03385.
[8] Zhang J H, Kong F T, Wu J Z, et al. Cotton disease identification model based on improved VGG convolution neural network[J]. Journal of China Agricultural University, 2018, 23(11):161-171.
[9] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015:770-778.