ITAIC 2020 (ISSN 2693-2865)

Image Style Transfer Method Based on Improved Style Loss Function

Hanmin Ye[1] , Wenjie Liu[1],Yingzhi Liu[1]
1. School of Information Science and Engineering Guilin University of technology Guilin, China
yehanmin@163.com,2529502268@qq.com, vsracer@hotmail.com

2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC) | 978-1-7281-5244-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ITAIC49862.2020.9338927

***Abstract*—In order to improve the quality of composite image in the process of image style transfer.This paper proposes an image style transfer method based on an improved style loss function: the improved Gram matrix calculates the inner product of the feature map and the spatial transformation map, and then calculates the new style loss function. At the same time, combined with the content loss function, the weighted algebraic sum of the two loss functions is used as the total loss function of the neural network. The gradient descent algorithm is used to iteratively optimize to generate the style-transferred image. Experimental results show that the Peak Signal to Noise Ratio and Structural Similarity values of this method are better than other style transfer algorithms, and the image texture details and spatial arrangement are more complete.**

Keywords—image style transfer, gram matrix, neural network, feature extraction

I. INTRODUCTION

As a new research direction in image processing, image style transfer has received extensive attention. Image style transfer Image style transfer, on the premise of ensuring the same semantic content of the original image, convert the style of different images, so as to achieve the purpose of the original image transfer style. But the image style is an abstract concept. Everyone has a different understanding of the style of different images or even the same image. How to use a computer to describe the style of an image more accurately is the difficulty of this research direction.

In the early stage, image style transfer was defined as texture synthesis [1], that is, while restricting the texture synthesis, the texture was synthesized from the source image to retain the semantic content of the source image. Efros [2] proposed a simple texture algorithm in which a new image is synthesized by stitching together small patches of existing images. Hertzmann [3] proposed an image analogies, which synthesizes images with new textures through image feature mapping relations. Ashikhmin [4] believes that in this process, the style that he wants to transfer is more important, so he focuses on transmitting high-frequency texture information while preserving the rough scale and semantics of the target image. Lee [5] controlled texture conversion through edge orientation information. Due to the limitations of scientific research at that time, cumbersome formulas and complex data models were needed to learn how to generate texture, and the establishment of mathematical models took time and effort, so the development of image style migration was very slow [6].

Since the introduction of deep learning, in the field of image processing, algorithms based on deep learning have successfully surpassed many traditional machine algorithms. Deep learning is used to extract image style features, and combined with the convergence of loss function, the research of image style transfer comes into people's view. Gatys [7] first proposed an image style transfer algorithm based on convolutional neural network. The core idea of this algorithm is to use the convolutional neural network to calculate the Gram matrix of the image, so as to obtain the image style features. This algorithm aroused the upsurge of image style transfer research. Johnson[8] and Ulyanov [9] used the feed-forward neural network for rapid style transfer. Zhu [10] uses GAN (Generative Adversarial Networks) network confrontation training model to increase the diversity of image conversion. Dumoulin[11] proposed a method that N styles share one model, which reduces the space consumed by the model and realizes real-time style interpolation on the video. Luan [12] added regular expressions to CycleGAN (Cycle Generative Adversarial Networks) to further improve the quality of generated graphics. Li[13] has greatly improved the color representation of style migration by adding whitening transform and coloring transform.

In this paper, the pre-trained VGG-19 model is used to obtain the content semantic information and style semantic information of the image, and a white noise image is used as the original input. The method of supervised learning is used to match the content semantic information and style semantic information of white noise images with the content semantic information and style semantic information of the content image respectively. Finally, use gradient descent to optimize. Compared with the Gatys [7] algorithm, this algorithm uses the improved Gram matrix to enhance the global arrangement information of the style features of the style image, make the generated image more stable, and reduce the style distortion and style dispersion problems.

II. IMAGE STYLE TRANSFFER BASED ON DEEP LEARNIN

A. Content Loss Function

Generally each layer in the VGG neural network defines a non-linear filter bank whose complexity increases with the position of the layer in the neural network. Given an input image p is encoded each layer of neural network. Each convolution layer l has N_l filters, and the size of each filter is M_l. The response in the l layer can be stored in the matrix $P_{i,j}^l \in R^{N_l \times M_l}$, $P_{i,j}^l$ is the response of the ith filter at the position j in the l layer.

In order to obtain the image information in each layer, the white noise image is used for gradient descent to find the image that matches the feature response of the original image. Let p and t be the content image and the composite image, and $P^l$and T^l be the feature response at the l layer. Define the content loss function as:

$$L_{content} = \frac{1}{2}\sum_{i,j}\left(T_{i,j}^l - P_{i,j}^l\right)^2 \qquad (1)$$

B. Style Loss Function

By using the Gram matrix to obtain the texture information of the input image, the feature correlation between the feature responses of different filters is stored in the Gram matrix. The Gram matrix of F_{ik}^l and F_{jk}^l of feature mapping at the l layer is:

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l \tag{2}$$

A stable multi-scale image representation is obtained by multi-layer feature correlation. It captures texture information of an image, but does not capture global texture information. The average square distance between the Gram matrix of the style image and the Gram matrix of the composite image is minimized by gradient descent of the white noise image. Let a and t be used as the style image and the composite image, A^l and G^l their respective style representation in layer l. The style loss of the l layer is:

$$E_{gl} = \frac{1}{4N_l^2 M_l^2} \sum_{ij} \left(G_{ij}^l - A_{ij}^l \right)^2 \tag{3}$$

The total loss style function is:

$$L_{style} = \sum_{i=0}^{L} \omega_l E_{gl} \tag{4}$$

ω_l is the weight of loss function of each layer style.

In the feature map, each value is the result of the convolution of a convolution kernel at a specific position, and each value in the Gram matrix represents the inner product of the two feature maps. If the values of the two dimensions of the two feature maps in the same dimension are smaller, the obtained value is smaller, otherwise, the obtained value is larger. Therefore, the Gram matrix can show the correlation between two features, such as which two features appear at the same time, and which two features trade off. Therefore, the Gram matrix can well represent the style of an image

III. IMPROVED GRAM MATRIX

Although Gram matrix can extract the global static features of the image, it ignores the global arrangement information of the image, so the generated texture is scattered and the distortion of the lines will occur. To solve this problem, an improved Gram matrix can be used. Unlike the Gram matrix that directly calculates the symbiosis between different feature maps, the improved Gram matrix calculates the inner product of a feature map F^l and the space conversion feature map T(F^l). T is an interspace transformation, which is used to calculate the similarity between local features and adjacent features. $T_{x,+\delta}$ means that Algorithm flow chartthe feature map is shifted to the right by δ pixels in the horizontal direction, that is, the original feature map removes the first δ column of pixels, and $T_{x,-\delta}$ shows the removed δ column of pixels, and the improved Gram matrix can be obtained：

$$G_{x,\delta,i,j}^l = \frac{1}{M_l} T_{x,+\delta}\left(F_i^l\right) T_{x,-\delta}\left(F_j^l\right) \tag{5}$$

The improved Gram matrix is actually the correlation between the features at (i, j) and the features at (i, j+δ) or (i, j-δ), so that the permutation information of the features is well preserved. Equation 5 only calculates the improved Gram in the x direction. In order to take into account the whole game, the improved Gram in the y direction should also be counted. Use $G_{x,\delta}^l$ and $G_{y,\delta}^l$, respectively to obtain the improved Gram loss:

$$E_{impgl} = \frac{1}{2}\left[\left(G_{x,\delta}^l - A_{x,\delta}^l \right)^2 + \left(G_{y,\delta}^l - A_{y,\delta}^l \right)^2 \right] \tag{6}$$

A_x^l and A_y^l respectively represent the improved Gram matrix in the x and Y directions of the output image

The improved total style loss function is

$$L_{impstyle} = \sum_{l=0}^{L} \omega_l E_{impgl} \tag{7}$$

ω_l is the weight of loss function of each layer style.

IV. IMAGE STYLE TRANSFER

The total style loss in this article is defined as follows:

$$L_{total} = aL_{content} + bL_{impstyle} \tag{8}$$

α is the weight of the content loss function, b is the weight of the style loss function,

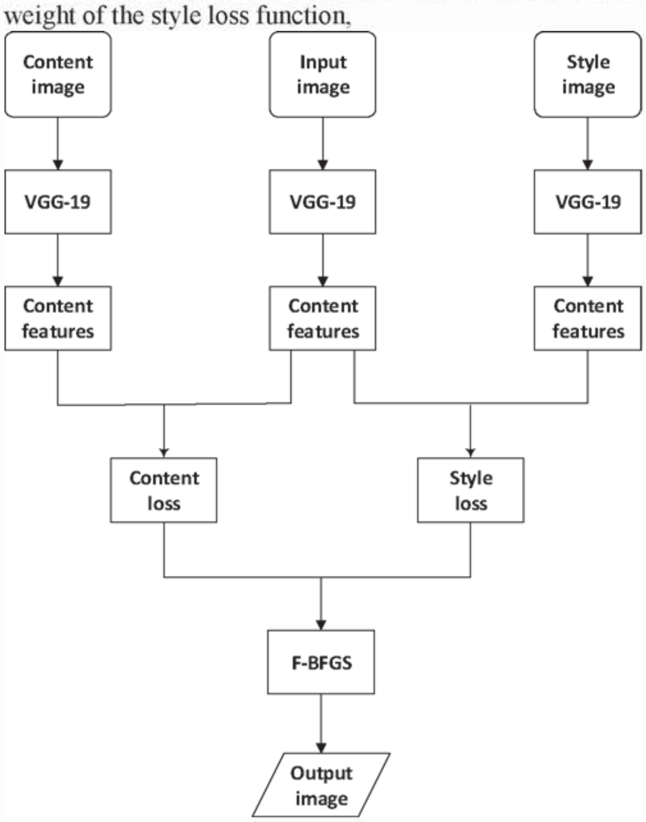


Fig. 1. Algorithm flow chart

The algorithm steps of this paper are as follows:

Step 1: Input content image and style image, and calculate content features and style features of each layer by VGG-19 network;

Step 2: Input random noise image and calculate its content features and style features respectively by VGG -19 network;

Step 3: Match the content features of the content image and the white noise image to calculate the content loss function; the style features of the style image and the white noise image want to match and calculate the style loss function;

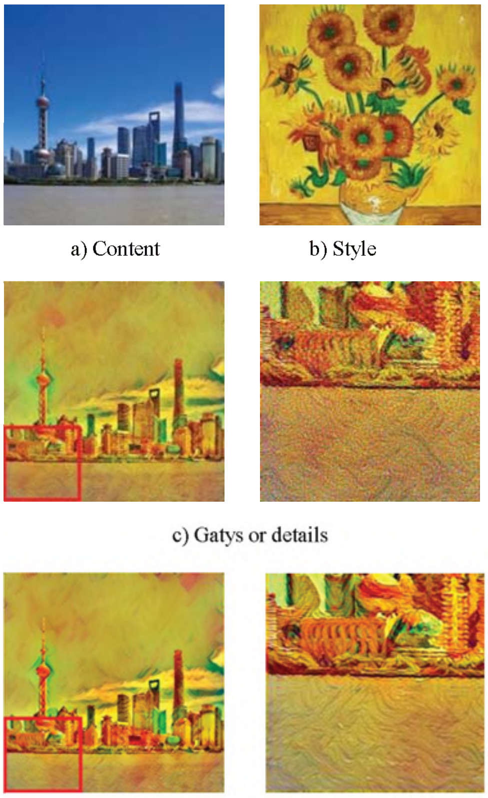
Step 4: Substitute the weighted algebraic sum of these two loss functions as the total loss function of the neural network into the neural network;

Step 5: Use F-BFGS algorithm to optimize the error function for operation;

Step 6: Output style transfer image

In this paper, the trained VGG19 model is used in the experiment, and the pooling method adopted is the homogenization method. The initialized output image is the original content image. The input content image needs to be preprocessed, that is, the mean RGB is subtracted. The content loss layer of conv4_2 and the improved Gram loss layer were pool1, pool2, poo31and pool4. The number of translation pixels is fixed at 4. In each loss function, the proportion of each layer is equal, and the proportion of style and content in the generated image is 100 and 1 respectively. The images presented in this paper were synthesized in a resolution of 256×256 pixel.

V. EXPERIMENTAL RESULTS



a) Content b) Style

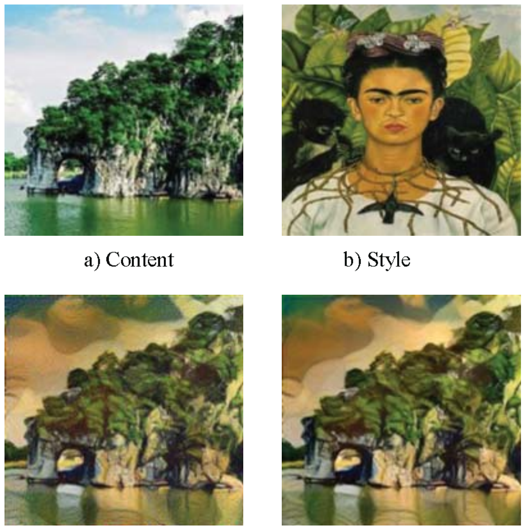
c) Gatys or details

d) New ways or details

Fig. 2. Comparision of result(Ⅰ)

. As shown in Fig 1, it can be found that the image quality obtained in this experiment has been greatly improved. As shown in the red box, there is a lot of high-frequency noise in the image synthesized by Gatys, there are many light or dark pixel particles, and the texture information of the style image is lost. The improved Gram matrix is used to calculate the similarity between local features and adjacent features, which makes the texture of style images smoother and thus improves the viewing effect of images

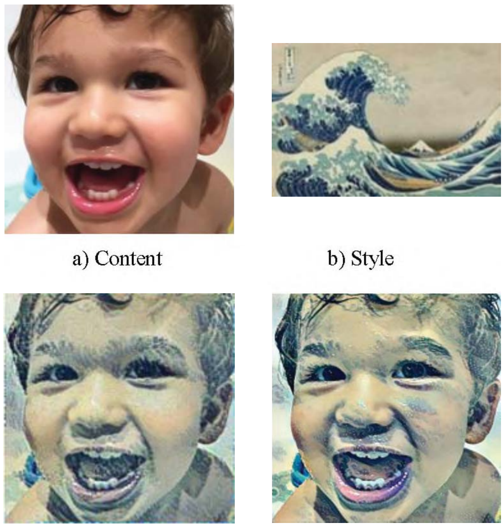
As shown in Fig 2, in the algorithm imaging effect of this paper, the leaf texture of the style image is clearer and the visual effect is more. As shown in Fig 3, the texture information of the style image transmitted by the algorithm in this paper is more silky. The characteristics of the style image transmitted by the Gatys algorithm have more texture spots, and the texture information is incomplete.



a) Content b) Style

c) Gatys d) New way

Fig. 3. Comparision of result(Ⅱ)



a) Content b) Style

c) Gatys d) New way

Fig. 4. Comparision of result(Ⅲ)

As shown in Table 1, the content image and style image are respectively PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity) operated with the composite image of Gatys algorithm and this algorithm. Different images have different style features and content features, which cannot be used as experimental basis. Only PSNR and SSIM can be compared for the same experiment. In the same experiment, PSNR value and SSIM value of this algorithm are always larger than Gatys algorithm, so the imaging quality of this algorithm is better.

TABLE I. COMPARISON OF PSNR AND SSIM RESULTS

		Fig 2 Experiment		Fig 3 Experiment		Fig 4 Experiment	
		PSNR	*SSIM*	*PSNR*	*SSIM*	*PSNR*	*SSIM*
Content	*Gatys*	6.3726	0.1700	9.6844	0.3411	0.3176	0.2418
	New ways	6.5589	0.5568	11.2927	0.5796	113.6278	0.5384
Style	*Gatys*	10.1151	0.0500	8.5161	0.0400	9.5998	0.0839
	New ways	12.1857	0.1689	9.2342	0.0880	9.9158	0.1606

VI. CONCLUSION

The image style transfer method based on the improved style loss function solves the problem by improving the operation rules of the style loss function. This algorithm improves the image quality while achieving image style transfer. When using neural networks for style transfer, the Gram matrix extracts the global static of the image, and the relationship between adjacent pixels of the same image is not effectively extracted. Through the improved Gram matrix, the similarity between local features and adjacent features is calculated. Solve the problem that the details of style characteristics are not clear and the image quality is poor. By calculating PSNR and SSIM, the experimental value of this algorithm is better than that of Gatys algorithm. Experimental results show that the algorithm improves the image quality while reducing the image style, enhances the texture of the style image, and reduces the style distortion.

REFERENCES

[1] HEEGER D J,BERGEN J R. Pyramid-based texture analysis/synthesis[C]//Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques. New York:ACM,1995:229-238.

[2] Efros A A, Freeman W T. Image quilting for texture synthesis and transfer[C]//Proc of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York:ACM Press, 2001:341-346.

[3] Hertzmann A, Jacobs C E, Oliver N, et al. Image analogies[C]//Proc of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York:ACM Press, 2001:327-340.

[4] Ashikhmin N. Fast texture transfer[J]. IEEE Computer Graphics and Applications,2003,23(4):38-43.

[5] Lee H,SEO S,RYOO S,et al. Directional texture transfer[C]//Proceedings of the 8th International Symposium on Non-photorealistic Animation and Rendering. New York:ACM,2010:43-48.

[6] Huang X,Belongie S.Arbitrary style transfer in real-time with adaptive instance normalization[C]//Proc of International Conference on Computer Science,2017:1510-1519

[7] Gatys L A, Ecker A S, Bethge M. Texture synthesis using convolutional neural networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems.Cambridge:MIT Press,2015:262-270.

[8] Ulyanov D,Lebedev V,Vedaldi A,et al.Texture networks:feed-forward synthesis of textures and stylized images[EB].eprint arXiv:1603.03417,2016.

[9] Ulyanov D,Lebedev V,Vedaldi A,et al.Texture networks:feed-forward synthesis of textures and stylized images[EB].eprint arXiv:1603.03417,2016.

[10] Zhu J Y,Park T,Isola P,et al.Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//IEEE International Conference on Computer Vision.IEEE Computer Society,2017:2242-2251.

[11] Dumoulin V,Shlens J,Kudlur M.A Learned Representation For Artistic Style[EB].eprint arXiv:1610.07629,2016.

[12] Luan F,Paris S,Shechtman E,et al.Deep photo style transfer[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition,2017:6997-7005.

[13] LI Y,FANG C,YANG J. Universal style transfer via feature transforms[C] // Proceedings of Advances in Neural Information Processing Systems,2017.