Part based Face Stylization via Multiple Generative Adversarial Networks

Wu Zhou[1]*, Xin Jin[1], Xingfan Zhu[1], Yiqing Rong[1], Shuai Cui[2]
[1] Beijing Electronic Science and Technology Institute, Beijing 100070, China
[2] University of California, Davis, Davis 95616, USA
*Corresponding author email: jinxin@besti.edu.cn

Abstract. In recent years, due to the improvement of scientific research methods and the wide-open source and acquisition of related data sets, face stylization has become a hot research field and application direction. There is a need to stylize face images in many applications, such as camera beauty, artistic photo processing, etc. However, most of the current schemes are not satisfactory, and the resultant image synthesis traces are obvious, and the effect is relatively monotonous. Based on the study of image features and style representation, this paper proposes a general-purpose face image style transfer whole process scheme. It can fill the gap in local style transfer of face images. Among the existing face stylization methods, the face stylization method is more complex, and the resulting obvious image synthesis trace along with the single effect. The project innovates the existing technology that can split the whole picture and implements the following six functions. Including the segmentation of specific portrait parts (hair), the skin buffing and whitening of the face, the defuzzification of the photos, the style transfer of the hair, the messy hair removal, and the implementation of the big eye effect. This study can realize the automatic style conversion of specific face images quickly and with high quality.

Keywords: computer vision, generative adversarial network, style transfer, semantic segmentation

1 Introduction

Face stylization is an important technical method in the field of computer vision, which is widely used in camera beauty, film production, and artistic photo processing. Through computer image processing technology, the face style transfer can well integrate the content of the original image with the style of the styled image, thus implementing the transfer from the original image to the styled image. However, most of the images obtained by the current solutions have obvious traces of synthesis and the effect is rather monotonous. So, the development and improvement of face image style transfer technology has significant scientific significance and application value.

Traditional face style migration methods mainly generate line drawings [1] or style drawings [2] based on face shape and facial contour information, and the images generated by these methods generally have no specific artistic style, only simple line drawings with some color rendering [3]. Some methods create a dataset by collecting images of different styles, selecting a reference sample, and then rendering the original image into a styled image with that style [4, 5]. For the latter, it can be divided into two methods. One is to directly split the image into pieces or to split the facial features, match it with the image pieces in the style dataset in some way, get the most similar image pieces, and then fuse them into a complete stylized image [6, 7]. Another method is to learn a high-dimensional hidden space by a deep learning model [8, 9], map the original image as well as the styled image to this space, and then parse and restore the style by the corresponding decoding network to realize the transfer from original image to styled image [10]. Among the methods as above, the deep learning method is suitable to be applied to face images and has good performance in implementing face stylization.

By generating a general adversarial network [11] for deep learning in image style, the workload can be reduced, and can produce rich effects. In some cases, it is difficult to obtain the paired dataset of the traditional method of generating adversarial network. Therefore, in order to avoid the limitation of the traditional generation adversarial network algorithm that requires paired data in image processing and improve the effectiveness of style transfer, this paper introduces a new convolutional neural network [12] to replace the original residual network in the process of network formation, and through the loss function composed of the same mapping loss and perception loss, the two can jointly measure the loss of style transfer. This improves the network characteristics and reduces the influence of samples in the network, thus improving the image quality after style transfer. In addition, the stability of the results is improved and the convergence speed is also increased.

The key work of the study involves the following contents:

- Based on the abstract expression of images in deep learning, this paper explores the correlation between image features and semantic content in generative adversarial network, and how to use image features to achieve style transfer.
- For some problems in image stylization, this paper improves the cyclic consistency network CycleGAN. After qualitative research and quantitative testing, it is shown that the improved cyclic consistency network has achieved the improvement of realism and diversity when transferring images.
- A specific style data set is designed and produced while maintaining the ID information of the original image and the details of the hair texture of the eyes.
- A full process design of face local style transfer is proposed, including fine segmentation of hair, facial features and human body, image super-resolution processing, hair removal processing using image patching technology, and finally enlarged eye beautification processing in the later stage.

2 METHODOLOGY

2.1 Improved structure of generative adversarial network CycleGAN

2.1.1 Problem analysis

Although the generation adversarial network does solve some problems of the generation model, it also has some enlightenment for the development of other methods. However, due to its incompleteness, it has caused some new problems when overcoming the existing problems. The greatest advantage of generating adversarial network is also the root of its biggest problem. Because the genetic algorithm uses anti-learning rules, the theory cannot determine the convergence of the model and the appearance of the balanced point. In the process of training, it is necessary to keep the balance and consistency of the two adversarial networks, otherwise it is difficult to achieve good training results. In practical applications, the synchronization of the two kinds of adversarial networks can not be controlled, resulting in unbalanced training process. In addition, as a training model based on neural network, generative adversarial network also faces the common problem of neural network modeling, poor interpretation ability. The most critical point is that although the samples generated by the generation adversarial network are diverse, there is a phenomenon of collapse model [13], which may produce complex samples with little difference for humans.

CycleGAN completes the cross mapping between two pixels X and Y through two generation units and two discrimination unit networks [14]. In essence, it is a ring network system composed of two mirror symmetric generative adversarial networks. In the model, two generating networks and discriminant networks are designed, which can be transformed into different types of images after training. However, because of the need for cyclic consistency in this process, a cyclic loss function is also set.

2.1.2 Improvement of CycleGAN

The generation unit of the original CycleGAN uses the residual network [15], which is connected by the encoder, converter and decoder after full convolution [16]. The residual network has great advantages in the application field of video recognition technology, especially in the application field of target detection. In the generation unit network of the traditional CycleGAN network, a nine-layer residual module is used for 256x256 size pictures.

This paper attempts to replace the residual network in the network generation unit with DenseNet module [17] and improve the CycleGAN by combining to maintain the original CycleGAN structure.

Our improved network & generator structure is shown in Figure 1.

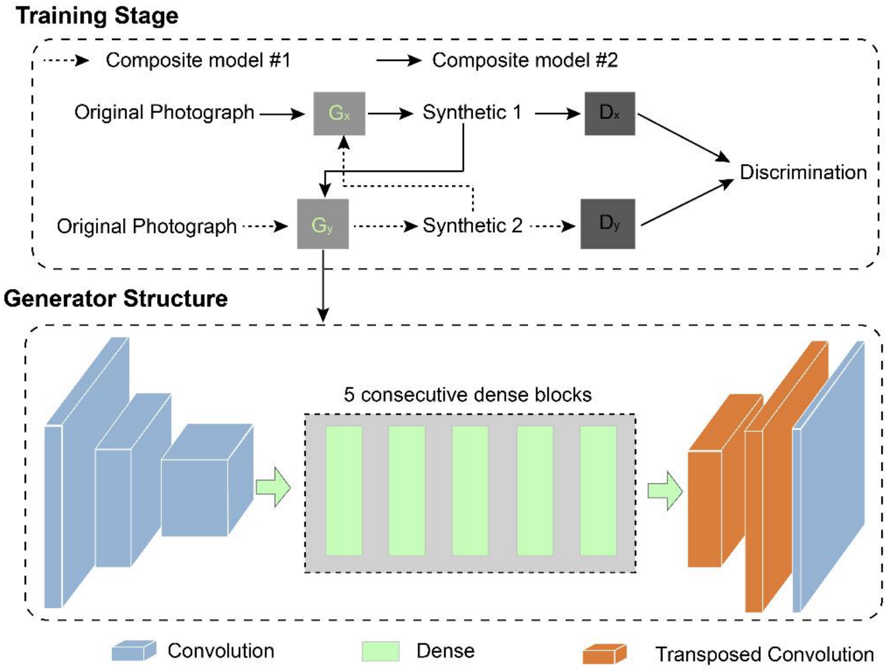

Fig. 1. Improved network & generator structure (DenseNet Module to replace ResNet Module)

2.2 Network design

In order to effectively avoid the technical limitations of the paired data required by the traditional generation adversarial network algorithm in the image style transfer process, and to improve the security and effectiveness of the image transfer process, this paper chooses to use an optimized and improved cyclic consistency reverse network system CycleGAN. DenseNet is used to replace the residual network in the network generation unit, and only a loss function composed of a mapping loss function and a perceptual loss function is used to calculate the loss reflecting style transfer. This idea greatly improves the network performance, effectively reduces the impact of network performance on paired data, and improves the image quality after style transfer. At the same time, the stability and convergence rate are improved.

(1) Encoder: Through convolution neural network, the characteristics are obtained from the input image.

(2) Converter: According to the different characteristics extracted, we can decide how to transform the feature vector of the image from the X domain (style photograph) to the Y domain (result photograph) The original CycleGAN converter uses 6-layer residual network blocks to transform the characteristic vector. Including nonlinear transformation function, input and output characteristics of residual network, normalization layer [18], convolution layer and ReLU layer [19].

(3) Decoder: Different from the decoding machine, the function of this module is to start the image from the feature vector value and gradually recover the underlying characteristics, so that the image can be generated. The implementation method is through the use of three anti convolution layers [20].

(4) Discrimination Unit: The discriminating unit of the improved network is PatchGAN [21] classifier. In the process of image discrimination and calculation by the image discrimination unit, the convolution between the two-dimensional input image block of the image and the input image block of the one-dimensional output image of the image is carried out layer by layer and point by point, and then the convolution and layer convolution of the one-dimensional output image block of each image are carried out to discriminate all the input output image blocks one by one by using the network. The arithmetic mean value of the judgment operation conclusion of the partition of each input image is taken as the final judgment result of the input image.

3 Detailed design scheme

This research involves converting the target image into a specific style while maintaining the ID information of the original image and the details of the hair texture of the eyes, including the data preprocessing module, the portrait segmentation network module, and the style transfer module. As is shown in Figure 2.

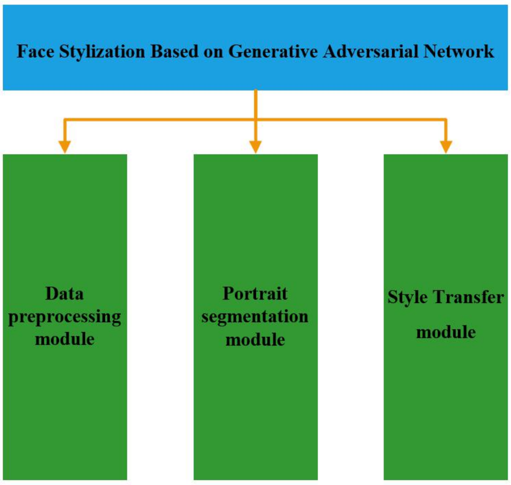
Face Stylization Based on Generative Adversarial Network
Data preprocessing module
Portrait segmentation module
Style Transfer module

Fig. 2. Module structure diagram of Face Stylization Based on Generative Adversarial Network

3.1 Data preprocessing module

Prepare face image dataset and target face style image dataset, wherein the face image dataset is from the public face image dataset provided on the network, and the target face style image dataset is from the specific style data set of Japanese big head post style. In order to better train the network in pairs, preprocessing such as image clipping and data enhancement is carried out for the network, including detecting the face and key points, correcting the face according to the rotation of the key points, expanding the boundary box of the key points in a fixed proportion and clipping the face area, and using the portrait segmentation model to set the background white.

3.2 Portrait segmentation module

Innovate the existing technology that can segment the whole picture and realize the segmentation of specific portrait parts (hair). First, the image of the preprocessed data set is input into bisenetv2. Secondly, the input image is represented by two branches (detail branch and semantic branch). Thirdly, through the enhanced training strategy like booster, the auxiliary segmentation header is inserted into different positions of the semantic branch, which further improves the segmentation accuracy of the image without increasing any inference cost. Again, two types of feature representation are enhanced by the designed guided aggregation layer. Finally, output the image after semantic segmentation. An example of segmentation training is shown in Figure 3.

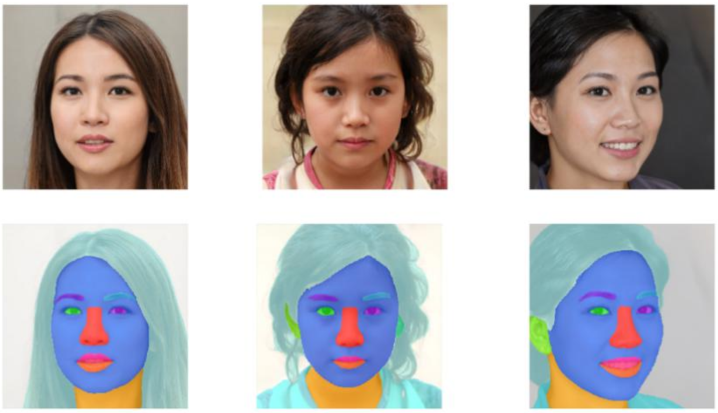

Fig. 3. An example of segmentation training

3.3 Style transfer module

Image style transfer is a subjective and exploratory design method. It is a method to design image changes according to the physiological characteristics of human visual system. However, most of the methods of degraded image restoration, including super-resolution image reconstruction, are based on an objective mechanism, so they can try to overcome it by trying to reproduce degraded images through reconstruction technology or using any prior knowledge in the process of image degradation, or by trying to use the completely opposite process of degraded images and images through image restoration technology or the technical model of image reconstruction. In addition, two technologies, excess hair removal and face authentication, are also applied.

It can be seen from Fig. 4 that the results of this process are good, the hair style is obviously transferred, the facial features are preserved, and the later beauty treatment is carried out.

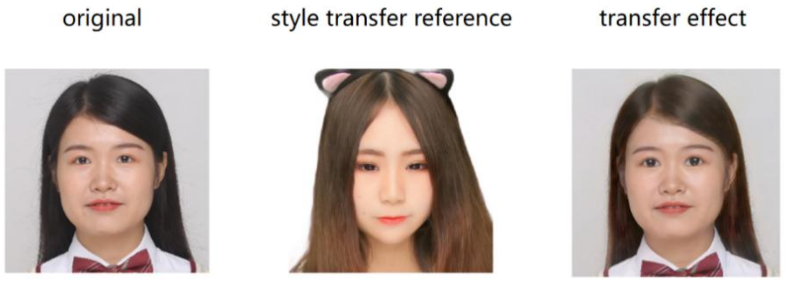

Fig. 4. Face beautification with style transfer processing effect

4 Experiment results

WGAN, CycleGAN and the modified CycleGAN were used to complete the image style transfer experiment independently. The results of the experiment are also shown in Figure 5 below. In Figure 5, the first column is the original drawing, the second column is the style drawing, the third column is the WGAN model result, the fourth column is the CycleGAN model result, and the fifth column is the improved CycleGAN result.

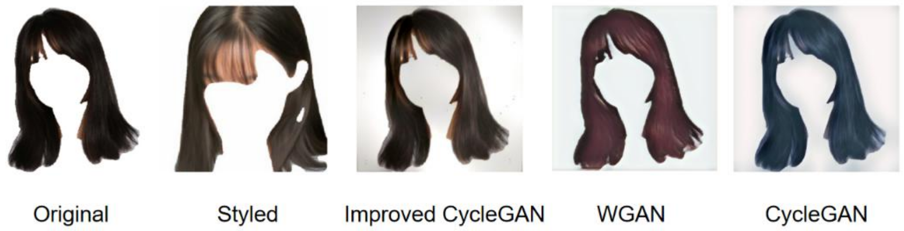

Fig. 5. Comparison results (Improved CycleGAN is our method)

Compared with the experimental results, we can further find that under the same number of iterations, all the schemes can complete the style transfer faster and better, and thus can obtain a more natural and real style transfer effect. The improved CycleGAN introduces a cyclic consistency loss function, which makes more effective use of the network bidirectional mapping model and prevents the collapse of the modeling itself to a certain extent. The improved CycleGAN provided in this paper also introduces the same perceptual loss function and mapping loss function. The flow of image style information transmission is relatively stable, but the modeling is not easy to collapse.

The whole process experiment of face local style is based on the Pytoch framework [22]. The experiment and test implementation are mainly divided into model part, training strategy part and evaluation index part.

It can be seen from Figure 6 that the effect of the whole process of face stylization is very good.

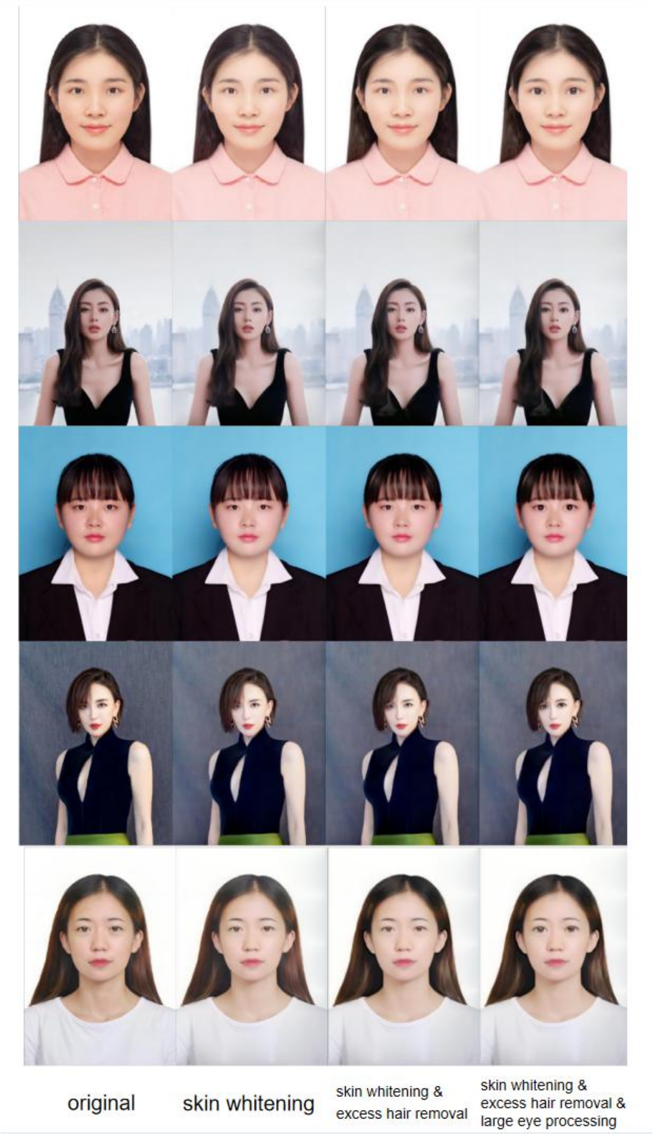
original
skin whitening
skin whitening & excess hair removal
skin whitening & excess hair removal & large eye processing

Fig. 6. The effect of the whole process of face stylization

5 Conclusion

This paper introduces semantic segmentation and combines it with other image processing technologies to solve the problem that local style transfer is not possible, which fills the gap in local style of face. Through the training experiment on the specific style dataset, a good local face style effect is obtained, which fully shows the performance of the model.

References

1. S. E. Brennan. Caricature generator: the dynamic exaggeration of faces by computer[J]. Leonardo, 1985, 18(3): 170-178.
2. Y. Li, H. Kobatake. Extraction of facial sketch based on morphological processing[C]. In: 1997 IEEE International Conference on Image Processing. 1997, 316-319.
3. M. Kaneko, M. Meguro. Synthesis of facial caricature using Eigenspaces and its applications to humanlike animated agents[C]. International Workshop of Lifelike Animated Agents –Tools, Affective Functions, and Applications, Tokyo, 2002, 58-63.
4. C. C. Tseng, J. J. J. Lien. Synthesis of exaggerative caricature with inter and intra correlations[M]. P P Lecture Notes in Computer Science. Heidelberg: Springer, 2007, 4843: 314-323.
5. S. E. Librande. Example-based character drawing[D]. Massachusetts Institute of Technology, 1992.
6. L. Liang, H. Chen, et al. Example-based Caricature Generation with Exaggeration[C]. 10th Pacific Conferenceon Computer Graphics and Applications, Beijing, 2003, 386-393.
7. X. Wang, X. Tang. Face photo-sketch synthesis and recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(11): 1955-1967.
8. J. F. Liu, Y. Q. Chen, W. Gao. Mapping Learning in Eigenspace for Harmonious Caricature Generation. 14th ACM International Conference on Multimedia, 2006, 683-686.
9. C. Zhang, G. Liu, Z. Wang. Cartoon Face Synthesis Based on Markov Network[C]. International Symposium on Intelligent Signal Processing and Communication Systems, 2010, 1-4.
10. H. Li, G. Liu, and K. N. Ngan. Guided Face Cartoon Synthesis[J]. IEEE Transactions on Multimedia, 2011, 13(6): 1230-1239
11. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.
12. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, UK: MIT Press, 2016
13. Goodfellow I. NIPS 2016 tutorial: generative adversarial networks. arXiv preprint arXiv: 1701.00160, 2016)
14. Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.
15. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
16. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

17. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
18. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International conference on machine learning. PMLR, 2015: 448-456.
19. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks[C]//Proceedings of the fourteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2011: 315-323.
20. Zeiler M D, Krishnan D, Taylor G W, et al. Deconvolutional networks[C]//2010 IEEE Computer Society Conference on computer vision and pattern recognition. IEEE, 2010: 2528-2535.
21. Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.
22. Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. Advances in neural information processing systems, 2019, 32.