Artificial Intelligence Technologies and Applications
C. Chen (Ed.)

doi:10.3233/FAIA231363

A Makeup Transfer Model for Half-Body Image

Yanting ZHAO[1], Guohua GENG[2], Long FENG[3], Haodong WU[4], Qihang LI[5]
School of Information Science and Technology, Northwest University, Xi 'an, China

Abstract. To apply makeup style of half-body reference images with complex makeup styles to the source image, we propose a model that combines object detection and makeup transfer. This model uses MTCNN to detect the target of Qinqiang Opera makeup image, and obtains the facial image with only makeup, so as to prevent the effect of makeup migration from being blocked by hair ornaments or hairstyles. And use the facial image obtained from object detection that only contains makeup as the reference image for the makeup transfer section, and use the CPM model to obtain the final result of makeup transfer in the makeup transfer section. In addition, we also collected the Qinqiang Opera character makeup dataset QinqMakeup. The Qinqiang Opera makeup migration method can not only enhance the entertainment of Qinqiang Opera, but also attract young groups and expand its audience. The ancient Qinqiang Opera art will be integrated with modern digital technology to achieve the inheritance and development of Qinqiang Opera, a Intangible cultural heritage.

Keywords. makeup transfer, object detection, complex makeup styles, half body image

1. Introduction

In recent years, people have conducted research on facial makeup in the field of computer vision. Given an arbitrary facial image and the desired makeup style reference image, the purpose of makeup transfer is to transfer the makeup style of the reference image to the target image while preserving the identity information of the target image. Makeup transfer requires extracting makeup styles from reference images, analyzing facial structures to correctly transfer makeup between misaligned faces. There are many factors to consider, including lighting, head posture, occlusion and facial expressions. The Generative model based on deep learning is an advanced method to solve this problem.

[1] Yanting ZHAO, School of Information Science and Technology, Northwest University, Xi 'an, China; e-mail: 1802228246@qq.com
[2] Corresponding author: Guohua GENG, School of Information Science and Technology, Northwest University, Xi 'an, China; e-mail: ghgeng@nwu.edu.cn
[3] Long FENG, School of Information Science and Technology, Northwest University, Xi 'an, China; e-mail: fenglong@stumail.nwu.edu.cn
[4] Haodong WU, School of Information Science and Technology, Northwest University, Xi 'an, China; e-mail: 207194494@qq.com
[5] Qihang LI, School of Information Science and Technology, Northwest University, Xi 'an, China; e-mail: 958984248@qq.com

BeautyGlow [1] and BeautyGAN [2] can achieve realistic makeup transfer images when reference images are with simple makeup style. However, they do not perform well in complex makeup that relies on shape, texture, and position. SCGAN [3] can successfully achieve makeup transfer when the reference image and target image pose angles are different, but the transfer effect for extreme makeup styles is still unsatisfactory. The CPM [4] proposed by Thao Nguyen et al. based on BeautyGAN can achieve makeup transfer for target images with different pose angles and extreme. makeup styles in reference images, and can achieve excellent results. However, these methods do not consider the quality problem of the reference image in the actual makeup migration. Especially in Qinqiang Opera, the actual makeup reference image of the role that we can obtain is often a bust image with headwear and clothing rather than a facial image with only makeup. The recently proposed SSAT [5] can solve the problem of occlusion, but it is not applicable to Qinqiang Opera makeup.

In this article, we propose a new deep learning model that combines object detection and makeup transfer. In order to achieve this goal, we use MTCNN[6] to detect the target of Qinqiang Opera makeup images, and obtain facial images with only makeup, to prevent the effect of makeup migration from being affected by hair ornaments or hairstyles. And use the facial image obtained from object detection that only contains makeup as the reference image for the makeup transfer section, and use the CPM model to obtain the final result of makeup transfer in the makeup transfer section.

2. Method

2.1. Network Structure

Figure 1 shows the network structure of this article. It mainly includes three modules: object detection, UV mapping, and makeup transfer module. Face detection takes a half body reference image as input and processes it as a face reference image through an MTCNN network. The UV mapping module maps the face reference image and the original image into UV images, so that the influence of facial expressions and postures can be ignored during subsequent makeup transfer. The makeup transfer module obtains the results in UV space and then maps them into a two-dimensional image through UV reflection, obtaining the original image with reference image makeup.

2.2. Object Detection Module

This part of the model uses the MTCNN model, which can be divided into P-Net[6], R-Net[6], and O-Net[6] three-layer network structures. This model mainly adopts three cascaded networks, using the idea of candidate boxes and classifiers for fast and efficient face detection. These three cascaded networks are P-Net that quickly generates candidate windows, R-Net that filters and selects high-precision candidate windows, and O-Net that generates the final bounding box and facial key points.

2.3. UV Mapping Module

We can use a pre-trained PRNet[7] model (denoted as UV) to extract the corresponding UV position map S and UV texture T from an input facial image I. We use rendering functions as following to reconstruct the input image from these UV representations:

$$S,T = UV(I) \text{ and } I := UV^{-1}(S,T) \tag{1}$$

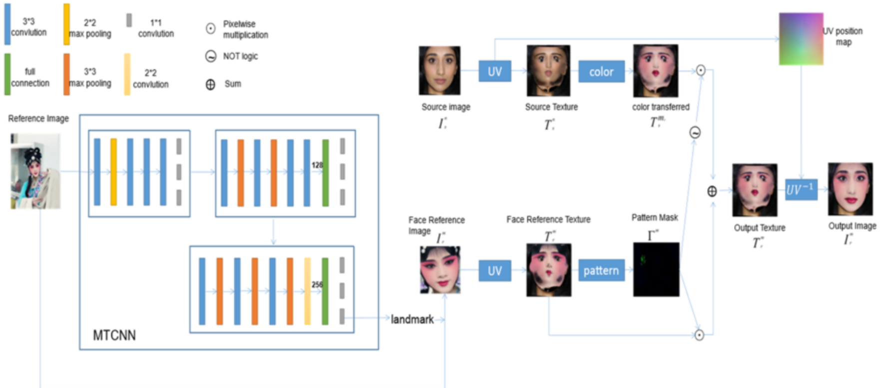

Figure 1. The overall network structure of the model

To Achieve makeup transfer in the case of spatial misalignment, we used these UV maps to illustrate. Firstly, we use UV converse function to get the corresponding UV mapping of each input image, which we call (S_s, T_s^n) and (S_r, T_r^m)

The UV position map of S_s and S_r only depends on the 3D face shape and is not related to the makeup style. Texture maps T_s^n and T_r^m are transferred to makeup transfer module to apply makeup of reference image to source image in UV space. The output of the makeup transfer is final texture image T_s^m. Finally, to convert this result texture image back to the standard image representation, we use the rendering function:

$$I_s^m = UV^{-1}(S_s, T_s^m) \tag{2}$$

2.4. Makeup Transfer Module

Color Transfer Module: This module uses a color-based makeup exchange network C to exchange makeup colors on the makeup area between the source image and the reference image. The architecture and training loss are proposed in BeautyGAN.

Pattern transfer module: This module aims to identify stickers, facial drawing, and decorative accessories, which we collectively refer to it as pattern-based makeup components, and tranfer them from reference image to the source image. When transferring these modes, we need to maintain their shape, texture, and position unchanged, but twist them onto the target 3D surface. In the form of natural images, this process is complex, including segmenting patterns, removing warping, and re warping on the target. Due to the representation of the UV position map, we do not need the steps of unwinding and rewinding. This problem can be simplified to simple image segmentation. Given the input texture mapping, our goal is to extract the composition pattern of a binary segmentation mask. We can achieve this by using any segmented network. In our implementation, a typical UNet[8] structure and a pre trained encoder Resnet-50[9] were used.

2.5. Loss functions

Loss functions of color transfer module: Loss functions of this module are the same of BeautyGAN. BeautyGAN use a network C to swap color of makeup on cosmetic regions between the reference image and source image.

$$T_s^{m_c}, T_r^{n_c} := C(T_s^n, T_r^m) \tag{3}$$

It uses four loss functions as following for training:

• **Adversarial Loss** [10]: L_{adv} uses two discriminators to enforce that the output maps T_s^{mc} and T_r^{nc} are in the non-makeup and makeup domains, respectively,

• **Cycle Consistency Loss** [11]: L_{cyc} applies the cycle consistency constraints that were proposed by CycleGAN,

• **Perceptual Loss** [12]: L_{per} use the VGG-16 model after training on ImageNet to restraint the identity of the source image and result image with makeup of reference image.

• **Histogram Matching Loss** [2]: L_{hist} intendes to distribute colors of the source images and the reference images consistent after makeup transfer.

The last loss function is the most critical loss function. It is Histogram Matching (HM) function proposed by BeautyGAN. It matches the histogram of the reference image and the source image on the eyeshadow, lips, skin and other areas to ensure the consistency of the source image and the reference image in the makeup color. The total loss is a weighted sum of the regional losses.

$$L_{hist} = \lambda^{eyes} L_{hist}^{eyes} + \lambda^{lips} L_{hist}^{lips} + \lambda^{skin} L_{hist}^{skin} \tag{4}$$

Where these hyper-parameters $\lambda^{eyes}, \lambda^{lips}, \lambda^{skin}$ are tunable.

Each loss term L_{hist}^i (i can be eyes, lips, or skin) is the distance between the histogram-matched version and the result image with makeup of reference image:

$$L_{hist}^i = \| T_s^{mc} \odot \Gamma_s^i - HM(T_s^n \odot \Gamma_s^i, T_r^m \odot \Gamma_r^i) \| \tag{5}$$

where $\odot$ denotes pixel-wise multiplication, Γ_r^i and Γ_s^i are the segmentation masks for region i in the source image and the reference image, respectively.

Loss functions of pattern transfer module: In this module, we use a pre-trained encoder Resnet-50 and a typical UNet structure in this module. For training, we use Dice Loss:

$$L_{DC} = \frac{2|\Gamma^{gt} \cap \Gamma^{pr}|}{|\Gamma^{gt}| + |\Gamma^{pr}|} \tag{6}$$

Γ^{gt} and Γ^{pr} refer to the ground truth and predictive segmentation masks of the pattern. We use the CPM-Synt-1 dataset to train this network, which includes annotated masks for the makeup patterns.

3. Experiment and Result

3.1. Experiment Detail

The model in this article combines object detection and makeup transfer. The makeup transfer module uses the CPM model, which is improved based on the BeautyGAN method for half body images. Therefore, we compared the proposed model with CPM and BeautyGAN, and used image quality evaluation indicators SSIM[13] (Structural Similarity Index) and PSNR[13] (Peak Signal to Noise Ratio) to evaluate it.

3.2. Dataset

QinqMakeup dataset. The MT dataset[14] contains a total of 3834 images, of which 1115 have no makeup and 2719 have makeup. We use the 1115 images without makeup, plus the makeup images of Qinqiang Opera characters we collected randomly select 100 unpasteurized images from them, and use our dataset QinqMakeup as the test set for makeup images. Use the images from our dataset QinqMakeup as reference images to perform makeup transfer on non makeup faces in the MT dataset, and check the makeup transfer ability of our model.

3.3. Experiment Result

Figure 2 shows our experimental results. The first row is the makeup image of our Qinqiang Opera characters, and the first column is the non-makeup source face image. The second line is the makeup transfer result image of BeautyGAN, the third line is the makeup transfer result image of CPM, and the fourth line is our makeup transfer result image.

Figure 2. Comparison of the Experimental Results of the Role Makeup Transfer in Qinqiang Opera

From Figure 2, we can find that our method has better visual effect than BeautyGAN and CPM in the role makeup migration task of Qinqiang Opera. It can be seen from the figure that BeautyGAN only transfered the makeup color of the reference image, and CPM can better transfer some makeup details, such as eye shadow, eyeliner and other makeup details, but it will be affected by problems such as hair occlusion in the image. Our method can eliminate the impact of hair style and hair accessories obstruction, achieving better results than CPM.

3.4. Image Quality Evaluation

Table 1 shows the comparison results of CPM and our proposed method using two image quality evaluation metrics SSIM and PSNR on the Qinqmakeup dataset. The data in the table is the PSNR and SSIM values obtained by randomly selecting 50 images from the dataset for calculation.

Table 1. The Quality Evaluation of the Experimental Results of Makeup Transfer

Method / **Dataset**	BeautyGAN		CPM		**Ours**	
	SSIM	PSNR	SSIM	PSNR	**SSIM**	**PSNR**
QinqMakeup	0.674	21.37	0.736	23.22	**0.844**	**27.42**

It can be seen that the method we proposed consistently achieves higher SSIM and PSNR scores on the QinqMakeup dataset. This indicates that our model has successfully completed the makeup migration of Qinqiang Opera roles and achieved better performance in the makeup migration task. Compared to the baseline method, the generated images exhibit superior quality, richer details, and more effective makeup transfer. Compared with BeautyGAN, the average SSIM of the model in this article increased by 25.22%, and the average PSNR increased by 28.31%. Compared with CPM, the average SSIM of the model in this article increased by 14.67%, and the average PSNR increased by 18.09%.

3.5. Model Applicability Evaluation

In order to further verify the universality of the model in this paper on the role makeup migration task of Qinqiang Opera, we show some results of Qinqiang Opera role makeup migration on the Qinqmakeup dataset. Figure 3 shows the experimental results

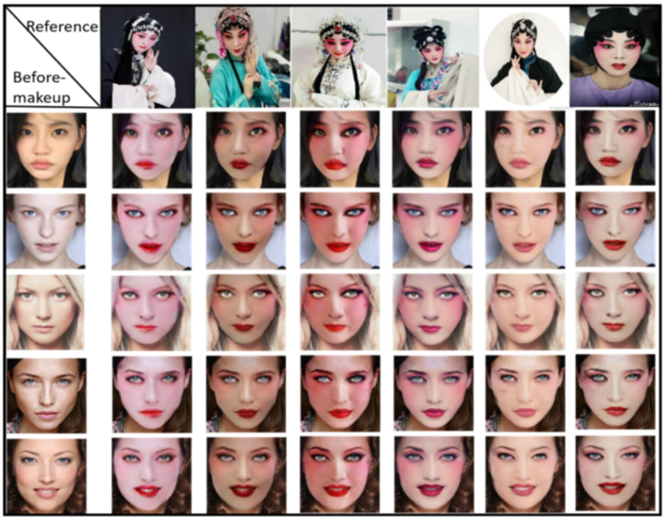

Figure 3. The experimental results of the generality of the model in this paper

From Figure 3, it can be seen that the model in this article can generate makeup transfer result images on different target images and under different reference images. In addition, when the reference images are the same, the proposed makeup transfer method performs well in makeup transfer tasks for different target images. The generated image exhibits a makeup style that preserves the identity features of the target image and high-quality transfers the reference image, consistent with human visual aesthetics. This shows that the proposed model has certain versatility and robustness in the role makeup migration task of Qinqiang Opera.

4. Conclusion

We propose a model that combines object detection and makeup transfer. This model uses MTCNN to detect the target of Qinqiang Opera makeup image, and obtains the facial image with only makeup, so as to prevent the effect of makeup migration from being blocked by hair ornaments or hairstyles. And use the facial image obtained from object detection that only contains makeup as the reference image for the makeup transfer section, and use the CPM model to obtain the final result of makeup transfer in the makeup transfer section. The experimental results show that the method in this paper can transfer the makeup style of Qinqiang Opera characters to the target image, retain the identity characteristics of the target image, and generate high-quality, real makeup migration result images. On the QinqMakeup dataset, the image quality evaluation index of the proposed method is superior to CPM, with an average improvement of 14.67% in SSIM and 18.09% in PSNR.

Acknowledgment

This research was funded by the the National Natural Science Foundation of China (62271393), National Key Research and Development Program of China (2020YFC1523301 and 2019YFC1521103).

References

[1] Chen, Hui, et al. "BeautyGlow: On-Demand Makeup Transfer Framework with Reversible Generative Network." 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019.

[2] Jing Liao, Yuan Yao, Lu Yuan, and Gang Hua. "BeautyGAN: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4880-4889. 2018.

[3] Deng, Y., Liu, X., Yang, J., & Tao, D. (2021). Spatially-Invariant Style-Codes Controlled Makeup Transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10751-10760).

[4] Nguyen, Thao, Anh Tran, and Minh Hoai. "Lipstick ain't enough: Beyond Color Matching for In-the-Wild Makeup Transfer." arXiv preprint arXiv:2104.01867 (2021).

[5] Li, Y., Zhang, Y., & Liu, Y. (2022). SSAT: A Symmetric Semantic-Aware Transformer Network for Makeup Transfer and Removal. arXiv preprint arXiv:2112.03631.

[6] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10):1499–1503, 2016.

[7] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. Joint 3d face reconstruction and dense alignment with position map regression network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5297–5305, 2018.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Springer, Cham, 2015.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[10] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [J]. Advances in neural information processing systems, 2014, 27: 2672-2680.

[11] 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.

[12] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//European conference on computer vision. Springer, Cham, 2016: 694-711.

[13] Hore, A., & Ziou, D. (2010). Image quality metrics: PSNR vs. SSIM. In 2010 20th international conference on pattern recognition (pp. 2366-2369). IEEE.

[14] Tingting Li, Ruihe Qian, Chao Dong, Si Liu, Qiong Yan,Wenwu Zhu, and Liang Lin. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In Proceedings of the ACM international conference on Multimedia, 2018.