Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Personalized Facial Makeup Transfer Based on Outline Correspondence[†]

Mengying Gao[1] | Pengjie Wang*[2]

[1]School of Computer Science and Engineering, Dalian Minzu University, Dalian, China
[2]School of Computer Science and Engineering, Dalian Minzu University, Dalian, China

Correspondence
*Pengjie Wang, No. 31, Jinshi Road, Jinshitan Street, Jinzhou District, Dalian, China. Email: pengjiewang@qq.com

Present Address
Funding information Natural Science Foundation of Liaoning Province, Grant/Award Number: 2023-MS-133; Dalian Minzu University, Grant/Award Numbers: 110222, 140140

Abstract

Most existing makeup transfer techniques focus on light makeup styles, which limits the task of makeup transfer to color manipulation issues such as eye shadow and lip gloss. However, the makeup in real life is diverse and personalized, not only the most basic foundation, eye makeup, but also the painted patterns on the face, jewelry decoration and other personalized makeup. Inspired by the painting steps of drawing the outline first and then coloring, we propose a makeup transfer network for personalized makeup, which realizes face makeup transfer by learning outline correspondence. Specifically, we propose the outline feature extraction module and outline loss that can promote outline correspondence. Our network can not only transfer daily light makeup, but also complete transfer for complex facial painting patterns. Experiments show that our method can obtain visually more accurate makeup transfer results. Quantitative and qualitative experimental results show that the method proposed in this paper achieves superior results in extreme makeup transfer compared to the state-of-the-art methods.

KEYWORDS:
Makeup transfer, Image to image translation, Semantic correspondence

1 | INTRODUCTION

Makeup transfer is an image translation task that focuses on human face. In the task of makeup transfer, the input contains a face image without makeup and a reference makeup image. Its goal is to synthesize a new image, which not only has the makeup style of the reference makeup image, but also retains the identity, expression, posture, background and other content information of the source image.

In recent years, Generative Adversarial Network (GAN) technology has been widely applied to various computer vision tasks due to its ability to produce high-quality realistic images. GAN-based facial makeup transfer methods [1,2,3,4,5] can significantly improve the makeup transfer effect. On the other hand, there are many kinds of makeup styles, including daily simple makeup, stage performance makeup and personalized extreme makeup. However, most of the above makeup transfer methods were proposed for the purpose of simple style makeup transfer, and those methods can not achieve good results for personalized extreme makeup.

There is also some previous work on makeup transfer that focus on extreme makeup transfer. For example, LADN [6]proposed to use multi-scale overlapping local discriminator and asymmetric loss to realize makeup transfer and removal. CPM [7] first

[†]This is an example for title footnote.

converts both the non-makeup image and the reference makeup image into UV texture maps. They then use makeup migration as a combination of colour transformation and pattern addition, and achieved good results on both simple and extreme makeups.

Among the above methods of extreme makeup transfer, as shown in Fig1, we found that the biggest problem is that the makeup pattern cannot be completely transferred. So how can we make the makeup pattern completely transferred? We observed that in real life, when people paint, they first outline the pattern, and then color the pattern. For example, someone draws a flower on paper. If you want to draw the same flower, then at the very basic we need to draw the exact same outline as his flower. Therefore, if we want to get a complete personalised makeup pattern, we must first get an outline that matches the reference makeup pattern. Through the advanced image to sketch technology [8], we can get the complete outline of the facial painting patterns in the reference makeup images.

In order to solve the problem of complex makeup pattern transfer, we propose a outline correspondence-based makeup transfer method. The outline correspondence between the non-makeup image and the reference makeup image is first established using the self-attention mechanism. Then we rely on this dense semantic correspondence to locate the color and texture information of the non-makeup image at the corresponding positions of the reference makeup image. So the generated image style finely matches the reference makeup image. Experiments have shown that our method is effective in transferring personalized extreme makeup, as shown in Fig3. In addition, our method can also successfully transfer daily light makeup, as shown in Fig4.

The main contribution of our paper is as follows:

- This paper is the first to propose the use of outline sketches for personalized makeup transfer.To this end, we design Affine Outline Feature Extraction (A-OFE) module to extract the outline features of non-makeup images, and Outline Feature Extraction (OFE) module to extract the outline features of reference makeup images.
- We propose a outline loss to constrain the distortion of the reference makeup features to the outline shape features of the non-makeup images, and improve the integrity of the transfer of the colored patterns of the reference makeup to the non-makeup images.
- Our method can not only transfer simple makeup, but also achieve excellent results in the task of transferring personalized extreme makeup. The experimental results show that we have achieved the most advanced quantitative and qualitative performance.

2 | RELATED WORK

2.1 | Semantic Correspondence

The depth feature extracted by convolution neural network is the expression of image high-level semantics. Early studies on semantic correspondence [9,10] focused on matching of manual features. Long et al. [11] first proposed to establish semantic correspondence by matching depth features extracted from a pre-trained classification model. Subsequent studies further improved the corresponding quality by adding additional labels [12,13,14,15,16,17], using a coarse-to-fine strategy [18]or retaining reliable sparse matching [19]. [18]proposed a new approach that allows direct visual attribute migration between two images, using high-level abstract features to establish a semantic correspondence between the contents of the two images. [20] By aligning images from different regions with the middle domain, a dense corresponding relationship is established in the middle domain, and the corresponding relationship is established to achieve cross-domain image translation. He et al. and Lee et al. [21,22] propose colorizing the target image according to the semantic correspondence between the target image and the instance image.

2.2 | Makeup Transfer

In recent years, there have been many approaches based on generative adversarial networks [1,2,3,4,5]. BeautyGAN [1]was one of the first methods to use GAN for makeup transfer, which uses a pixel-level histogram loss to achieve instance-level makeup transfer. PSGAN [2] utilizes the face analysis map and facial coordinate points to construct the pixel-level correspondence between the source image and the reference image, so as to solve the problem of face misalignment between different head poses and facial expressions. DMT [23] is the first model to solve the makeup transfer problem by disentangling the representation. Compared to the previous model, this model has further improved the effect of makeup transfer, but still has some limitations and is only suitable for transferring simple makeup. SSAT [24] used semantic symmetric learning for makeup transfer and removal. LADN [6]

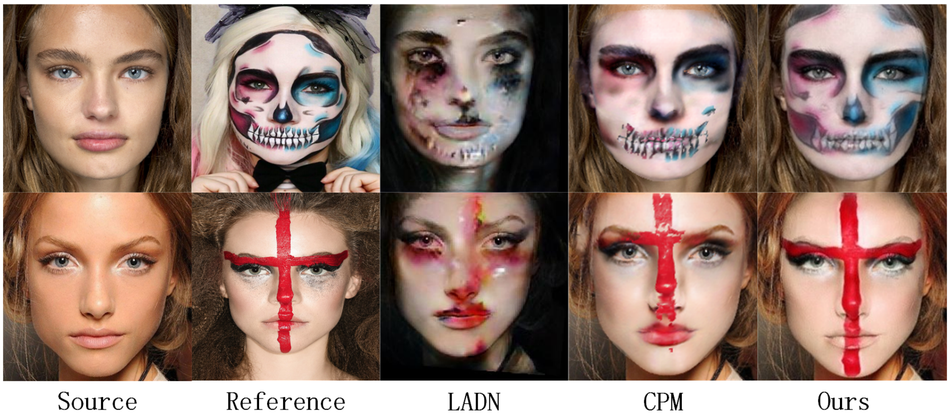

Figure 1 The result comparison of our method (Ours), with previous extreme makeup transfer methods, LADN and CPM.

uses multiple overlapping local discriminators and asymmetric loss functions to achieve extreme makeup transfer and removal. CPM[7] borrows the idea of PRNet[25] to implement makeup transformation in UV space. The color transfer branch and pattern transfer branch of it can be run independently, but the shortcoming is that the transfer pattern is not complete enough and the skin tones are not equal.

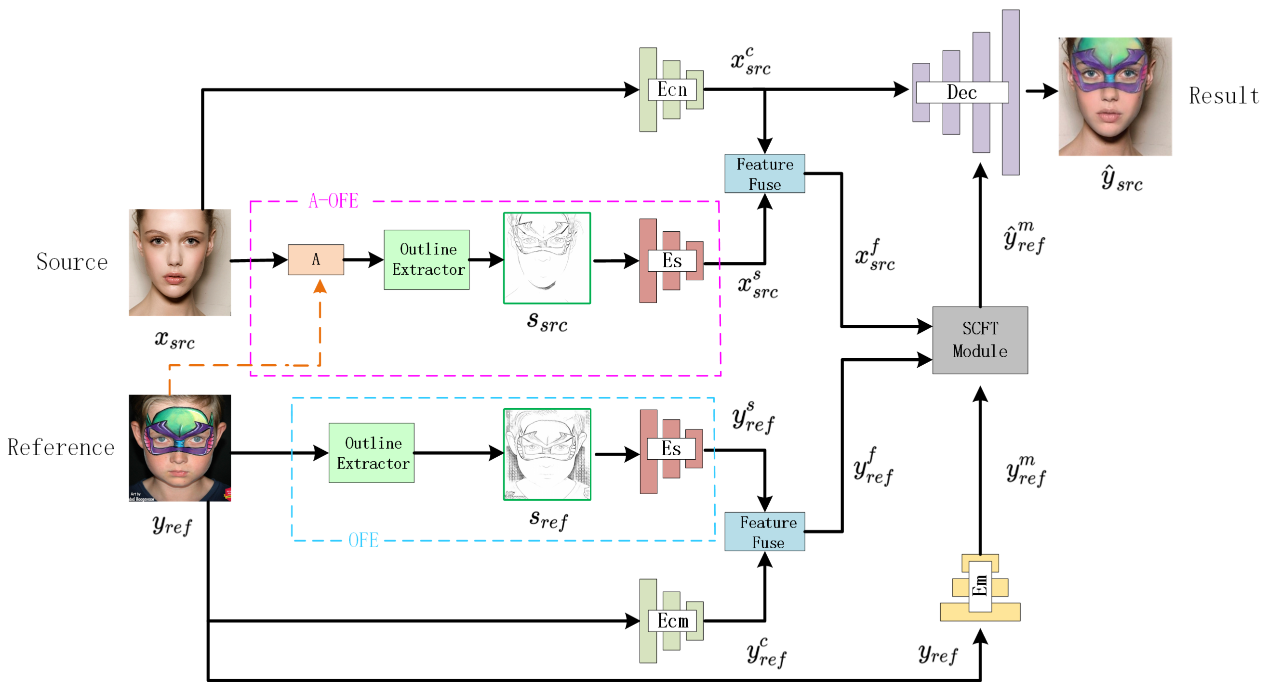

Figure 2 Personalized makeup transfer overall network framework. The input of the network is non-makeup images and reference makeup images, A-OFE (Affine Outline Feature Extraction) and OFE (Outline Feature Extraction) are both outline feature extraction modules. the input images need to go through Affine Transformation module A and Outline Extractor to obtain the required outline sketches, then each image is sent to Content Encoder E_{cn}, E_{cm}, Sketch Encoder E_s and Style Encoder E_m respectively. The Feature Fuse module[24] fuses the content features and outline features, and then the fused features are sent to SCFT (Spatially Corresponding Feature Transfer)[22] .module to establish the outline correspondence, and finally the decoder generates the makeup transfer result.

3 | PROPOSED METHOD

3.1 | Problem Formulation

The goal of our algorithm is to transfer the makeup style from a reference makeup image to a non-makeup image. In this paper, $X \subseteq R^{H\times W\times 3}$ refers to source domain with non-makeup images and $Y \subseteq R^{H\times W\times 3}$ refers to reference domain with makeup images. Given a non-makeup image $x_{src} \in X$ and a reference makeup image $y_{ref} \in Y$, the goal of makeup transfer is to learn a mapping function : $\Phi : x_{src}, y_{ref} \rightarrow \hat{y}_{src}$, where $\hat{y}_{src}$ has the same makeup style as y_{ref} while retaining the content features of x_{src}.

3.2 | Network Architecture

Fig2 describes our proposed network architecture for personalized makeup transfer. It mainly includes four encoders, namely content encoder E_{cn}, E_{cm}, reference image makeup style encoder E_m, and sketch encoder E_s. It also includes two outline feature extraction modules (A-OFE,OFE), feature fusion module(Feature fuse), spatial correspondence feature transfer module (SCFT), and a makeup decoderD_{ec}. The functions of each module are described in detail below.

Encoder

In this paper, we focus on personalised makeup transfers, which indicates that the reference makeup images differs significantly from the non-makeup image. To enable the network to accurately distinguish between the two domains, we use two different content encoders, E_{cn}, E_{cm}, to extract the content features of the non-makeup image $x_{src}$and the reference makeup imagey_{ref} respectively:

$$x^c_{src} = E_{cn}(x_{src}), y^c_{ref} = E_{cm}(y_{ref}) \tag{1}$$

In order to completely transfer the facial painted patterns and establish the outline correspondence between the reference makeup image and the non-makeup image, We use the sketch encoder E_s to extract the outline features of the image.

$$x^s_{src} = E_s(s_{src}), y^s_{ref} = E_s(s_{ref}) \tag{2}$$

To separate makeup style features from other features unrelated to makeup. We use a style encoder E_m to extract makeup style features from reference makeup images:

$$y^m_{ref} = E_m(y_{ref}) \tag{3}$$

Outline Feature Extraction

Two different modules have been designed to extract the outlines of the non-makeup image and the reference makeup image. The A-OFE (Affine-Outline Feature Extraction) is used to extract the outline features of the non-makeup image. In order to include makeup patterns in the outline sketch of the non-makeup image, we first need to obtain the corresponding pseudo-paired image. The specific process of generating pseudo-paired data and extracting outlines is shown in FigA1, which is detailed in the appendix. Then the outline sketch s_{src} is sent to the sketch encoder E_s for feature extraction. The module is shown in Fig2. OFE (Outline Feature Extraction) is used to extract the sketched outline features of the reference makeup image, here no affine transformation is applied to it, the image is fed directly into the outline extractor to obtain the corresponding outline sketch s_{ref}, and then the sketched outline features are obtained, see Fig2 for details of the module.

Outline Correspondence

The second step of the proposed algorithm is to fuse the content features and outline features of the image using the Feature Fuse method proposed in[24] to obtain richer features for feature matching. The process of feature fusion can be expressed as follows:

$$x^f_{src} = FeatureFuse\left(x^c_{src}, x^s_{src}\right) \tag{4}$$

$$y^f_{ref} = FeatureFuse\left(y^c_{ref}, y^s_{ref}\right) \tag{5}$$

We use the SCFT (Spatially Corresponding Feature Transfer) module in[22] to establish a dense semantic correspondence between outlines. The semantic correlation matrix obtained is computed to warp the reference makeup style features so that the makeup pattern outlines are spatially aligned with the outlines of the non-makeup image. This process can be expressed as:

$$\hat{y}^m_{ref} = SCFT\left(x^f_{src}, y^f_{ref}, y^m_{ref}\right) \tag{6}$$

Figure 3 Our personalized makeup transfer results.

Then, the makeup style features correspond to the outline of the non-makeup image in spatial position, and the color is distributed according to the shape of the outline. The facial makeup pattern of the reference makeup image can be accurately mapped to the non-makeup image.

Decoder

we put the content features of the non-makeup image into the decoder along with the distorted makeup style features to obtain the final makeup transfer result. We use the conditional normalization method proposed in [26], the spatially adaptive normalization block (SPADE) to project spatially varying makeup features to different activation locations. Embedding distorted makeup style features into content features for makeup transfer.This process can be expressed as:

$$\hat{y}_{src} = Dec\left(x^{c}_{src}, \hat{y}^{m}_{ref}\right) \tag{7}$$

3.3 | Objective Functions

Outline Loss

The gradient profile loss (GPloss) proposed in [27] can effectively extract shape features from the colour content of an image and calculate their similarity in that vector space by using the spatial outlines of the image as row vectors and column vectors, and then obtain the similarity of row outlines and column outlines of two images. In order to learn the correct outline correspondence, we calculate the outline similarity of the two sketches according to the above method and propose a outline loss that enhances the outline correspondence, in the way that the loss constrains the learning of the semantic correlation matrix so that the reference makeup outline can correspond to the non-makeup outline more accurately.

$$L_{outline} = GP\left(s_{src}, \hat{s}_{ref}\right) \tag{8}$$

$\hat{s}_{ref}$ is the result of the outline sketch s_{ref} after distorting according to the learned semantic correlation matrix. The constraining of the similarity between two outline sketches will implicitly push the learning of the semantic correlation matrix of fused features in the network.

Makeup Loss

Here we introduce Spatial Profile Loss (SPL) [27]to guide makeup transfer. Spatial Profile Loss consists of two losses, Gradient Profile (GP) loss [27] is used to constrain the shape of the image, and Colour Profile (CP) loss [27] refers to the loss in the gradient

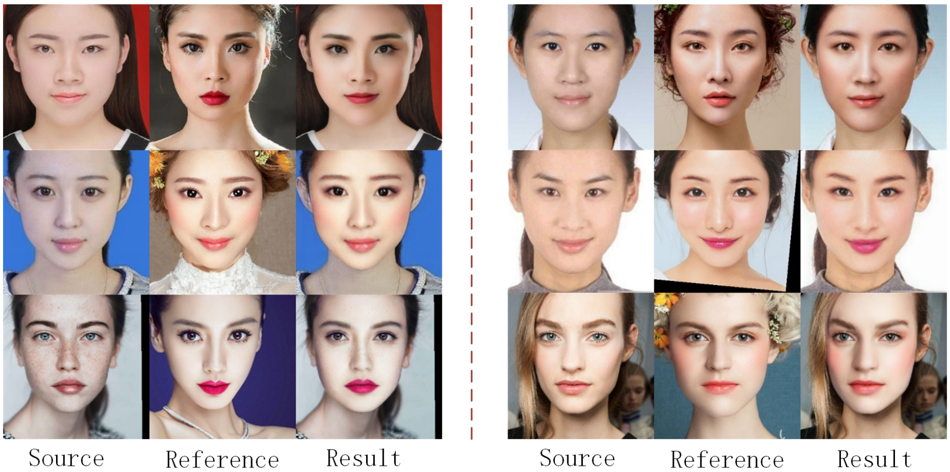

Figure 4 Makeup transfer results of our method on MT[1] dataset.

space of the computed chromaticity channel, which helps the network to learn stronger color transfer.

$$L_{makeup} = SPL\left(\hat{y}_{src}, x_{src}, \bar{y}_{src}\right) \tag{9}$$

Therefore, we use SPL to constrain the gradient consistency between $\hat{y}_{src}$ and x_{src} so that the generated results better preserve the person identity. The makeup style transfer is guided by constraining the color consistency between the makeup transfer result $\hat{y}_{src}$ and the pseudo-paired data $\bar{y}_{src}$.

Reconstruction Loss

We send the content and style features of the reference makeup image to the decoder to get the reconstructed reference makeup image. By using L1 loss between the reference makeup image and the reconstructed image, the encoder can correctly extract the makeup style features of the reference image.

$$y^r_{ref} = Dec\left(x^c_{ref}, y^m_{ref}\right) \tag{10}$$

$$L_{rec} = \left\|y^r_{ref} - y_{ref}\right\|_1 \tag{11}$$

Adversarial Loss

The discriminator D_Y is used in the makeup domain with the purpose of distinguishing the real samples from the generated ones and helping the generator to produce high-quality output. Least squares loss[28] is used for stabilization training.

$$L_{adv} = E_{y_{ref}}\left[\left(D_Y\left(y_{ref}\right)\right)^2\right] + E_{\hat{y}_{src}}\left[\left(1 - D_Y\left(\hat{y}_{src}\right)\right)^2\right] \tag{12}$$

Final Loss

In total, four loss functions are used for the training of the makeup transfer network. The overall loss is as follows:

$$\begin{aligned} L_{total} =& \lambda_{outline} L_{outline} + \lambda_{makeup} L_{makeup} \\ &+ \lambda_{rec} L_{rec} + \lambda_{adv} L_{adv} \end{aligned} \tag{13}$$

4 | EXPERIMENTS

4.1 | Dataset

CPM[7] offers the extreme style dataset CPM-Real, a dataset of real faces with a diverse range of makeup styles. These styles can range from light to heavy, including pattern driven style and colour oriented style. Many images have extreme makeup, including face decorations, face paint and holiday makeup. A total of 300 images of extreme makeup styles were selected from

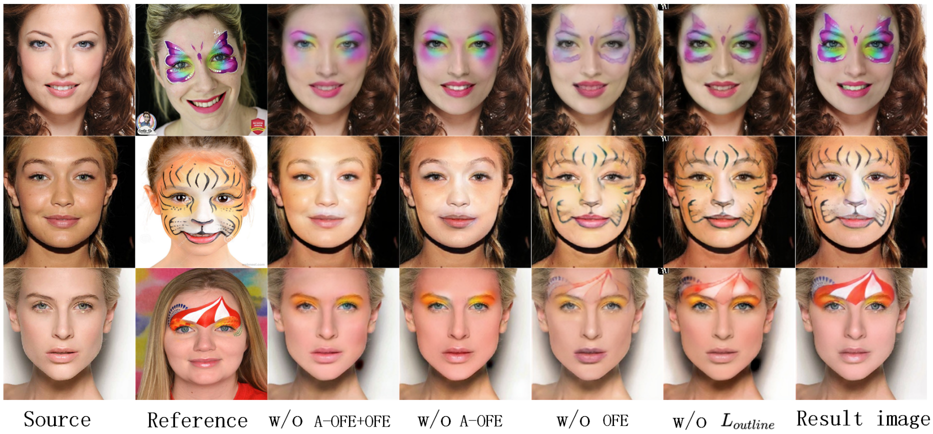

Figure 5 Ablation experiment.

the extreme makeup images collected by LADN[6] and CPM-Real. A total of 300 non-makeup images were randomly selected from the MT[1] makeup dataset. A total of 180,000 pairs of pseudo-paired data were generated for personality makeup transfer using the pseudo-paired image generation method proposed above. In addition to training on the CPM[7] dataset, we used the above method to train on the MT[1] dataset as well, and experimentally demonstrated that our algorithm is still valid for the simple makeup transfer task.

4.2 | Implementation Details

In the training, we scaled the size of the input image to 256×256, initialized all trainable parameters normally, and used an adam optimizer with $\beta_1 = 0.5$, $\beta_2 = 0.999$ for training. To balance the different loss functions, we set $\lambda_{Line} = 1.5$, $\lambda_{makeup} = 2$, $\lambda_{rec} = 1$, $\lambda_{adv} = 5$. Our experiments use the pytorch framework, and the batch size is set to 1 due to GPU memory limitations. 900 epochs are first trained with a learning rate of 0.0002, and the learning rate decays linearly to 0 for the next 500 epochs.

4.3 | Effect of Outline Loss

To verify the effect of the outline loss on the makeup transfer effect, we remove the loss and use the remaining loss for training, and the training results are shown in Fig5. In the case that the outline sketch is used without outline loss, although the pattern on the generated result looks the same as the pattern on the reference makeup image, some details of the pattern cannot be generated and there are some color differences in the pattern. After using outline loss, the pattern outline of the generated result is more accurate and clearer, and the color of the facial pattern is closer to that of the reference makeup image. The quantitative test results are shown in Table1, which also indicate that using outline loss will have higher similarity.

4.4 | Effect of Outline Sketches

The purpose of this paper is to establish the correct outline correspondence to improve the integrity of the transfer pattern. We remove the A-OFE module and the OFE module respectively to analyze the role of outline sketching in the makeup transfer task. The experimental results are shown in Fig5. After removing the A-OFE module, the generated result is similar to the reference image only in color, and cannot generate a specific pattern outline. After removing the OFE module, although the generated results can see a clear outline of the reference makeup, they cannot generate color features consistent with the reference makeup. If both modules are removed together, without using any outline sketches, the network cannot learn any outline correspondence. The painted pattern of the reference makeup cannot be transferred, and the face of the generated image has only fragmented colors. The results of the quantitative experiments are shown in Table1. It can be seen that all three metrics are influenced by the number of outline sketches.

4.5 | Qualitative Comparison

Here we have chosen to compare light makeup in everyday life with extreme styles of individual makeup looks. The qualitative comparison results are shown in Fig6. BeautyGAN[1] transfers well on many makeups, but fails to transfer individual eye shadows and face painting patterns. DMT[23] has a good transfer on light makeup, but fails to transfer extreme styles of face paint patterns. PSGAN[2] is also only able to transfer light makeup but not complex makeup. LADN[6] is able to transfer extreme makeup, but has problems such as blurring of the generated images and unclear makeup patterns. CPM[7] is very effective in transferring complex makeup, but still has certain problems, such as incomplete makeup patterns and uneven facial skin tones. Compared with these previous works, our method is state-of-the-art in transferring extremely complex makeup. The results are shown in Fig6.

Figure 6 Comparison with state-of-the-art methods. The first two rows are light makeup transfers and the last three rows are complex makeup transfers.

4.6 | Quantitative Comparison

We randomly selected 600 images generated from 15 non-makeup images and 40 reference images of extreme makeup styles for qualitative comparison. For a specific makeup style, the better makeup transfer method should generate images with a more similar distribution to its input reference images. Therefore, we measure the effect of makeup style transfer by calculating the Fréchet Inception Distance (FID) between the generated images and the pseudo-paired images of the makeup transfer method in recent years[29]. After makeup transfer, the generated result image should have a similar structure to the pseudo-paired image, so we use the Structural Similarity Index Metric (SSIM)[30] to evaluate the similarity between the two images. PSNR was used to test the quality of the generated images. The average scores of each metric are reported in Table2.

4.7 | User Study

To further measure the effectiveness of our method, we conducted a user study among 45 volunteers. We randomly selected 5 non-makeup images and 10 makeup images from each of MT dataset[1] and CPM-Real dataset[7] for our experiments. The images were fed into our model, and the models of BeautyGAN[1], DMT[23], PSGAN[2], LADN[6], and CPM[7], respectively. We divided the makeup migration task into light makeup transfer and complex makeup transfer. Each volunteer was presented with a sample of images generated by different models for each type of task and asked to select the best set of samples from multiple sets based on image quality and style consistency. A total of 45 questionnaires were collected, each containing the best model chosen by the volunteer for each type of task. Table3shows the percentage of each model selected, indicating that our method has the highest percentage of selection in complex makeup transfer tasks, but the percentage in light makeup transfer tasks is not high.

5 | CONCLUSION

We propose a personalized face makeup transfer method based on outline correspondence, by establishing a dense outline correspondence between the non-makeup image and the reference makeup image, relying on such correspondence to correlate the outlines of the two images and guide the generation of pattern outline shapes, and also according to the correspondence to enable the pattern colors of the reference makeup image to be accurately mapped to the non-makeup image. Experimental results show that our proposed method significantly outperforms other methods in terms of both image quality and integrity of painted patterns in the task of extreme makeup transfer.

Our method is mainly for complex makeup transfers, so it has some limitations for light makeup and does not retain the identity of the person in the target image well. In addition, it is unable to produce ideal makeup transfer results for facial images with very different poses and expressions, and the failed cases are shown in FigA3 in the Appendix. In future studies, we will focus on the case where the pose and expression of the source and target images are very different, and the preservation of the person identity of the target image.

ACKNOWLEDGEMENTS

We thank the Natural Science Foundation of Liaoning Province and the university for their support and assistance. The method proposed in this paper is for academic research only and cannot be used in any commercial field for legal profit.

ORCID

Pengjie Wang https://orcid.org/0000-0003-1276-9776

Table 1 Quantitative comparison results of ablation experiments.

A-OFE	**OFE**	$L_{outline}$	**FID↓**	**PSNR↑**	**SSIM↑**
			100.18	20.35	0.75
	√		84.91	21.07	0.76
√			75.71	20.64	0.77
√	√		63.77	21.34	0.77
√	√	√	**35.58**	**23.92**	**0.83**

□

Table 2 Quantitative experimental results.

	FID↓	PSNR↑	SSIM↑
BeautyGAN	129.64	17.02	0.67
DMT	123.96	17.73	0.73
PSGAN	132.35	17.27	0.63
LADN	110.90	9.07	0.28
CPM	50.32	11.35	0.35
Ours	**35.58**	**23.92**	**0.83**

Table 3 The ratio selected as best.

Makeup transfer tasks	BeautyGAN	DMT	PSGAN	LADN	CPM	Ours
Light makeup	20%	20%	22%	9%	11%	18%
Complex makeup	0%	0%	0%	13%	27%	**60** %

APPENDIX

A SUPPLEMENTARY MATERIALS

A.1 Pseudo-paired data generation and outline extraction

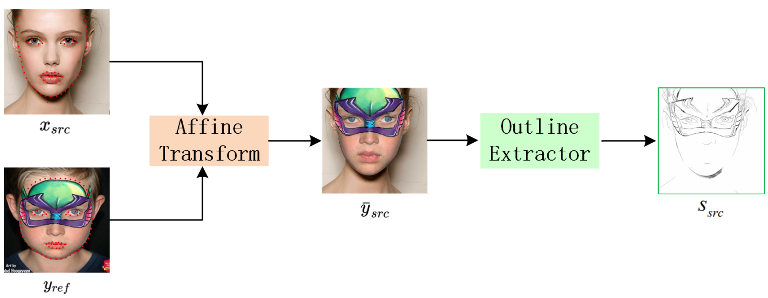

Figure A1 Obtaining pseudo-paired image and outline extracting.

In order to obtain a pattern shape that is consistent with the reference makeup without losing the original outlines of the characters in the non-makeup images, We manually selected 88 facial key points through face dense key point detection technology Face ++ API[1]. Affine transformation is performed on the facial key points of the non-makeup image and the makeup reference image, so that the makeup face is deformed accordingly. Then paste the deformed face with makeup on the non-makeup to get a pseudo-paired image $\bar{y}_{src}$. The outline extractor here makes reference to the image to sketch technique[8]. The process is shown in FigA1. At the same time, the pseudo paired data will be used as our ground truth in network training.

A.2 Network Structure

For network architecture of our encoders and decoder, we follow Sun et al.[24] networks. Specifically, all the encoders we use are composed of convolutional networks, including a convolutional layer, an instance normalization layer, and a ReLU activation

[1]Face ++ API: https://www.faceplusplus.com.cn/dense-facial-landmarks/

layer. The network architectures of Encoder E_{cn}, E_m are shown in TableA1. The network architecture of Encoder E_s is the same as that of E_m, except that the number of input channels becomes 1. In the decoder, we used the SPADE residual block to embed the makeup features into the fused features, and the network architecture is shown in TableA1. The specific network structure parameters are given in TableA1. N: the number of output channels, K: kernel size, P: padding size, S: stride size, IN: instance normalization.

Table A1 The network architecture of Encoder and Decoder. [24]

Layer	Encoder	Decoder
L1	Conv(N:64, K:7x7, S1, P3), IN, Leaky ReLU	Upsample:2, SPADE, Resnet
L2	Conv(N:128, K:3x3, S2, P1), IN, Leaky ReLU	Upsample:2, SPADE, Resnet
L3	Conv(N:256, K:3x3, S2, P1), IN, Leaky ReLU	SPADE, Resnet
L4	N/A	Conv(N:3, K:7x7, S1, 3), tanh

A.3 Makeup Style Interpolation

Makeup style interpolation means that we transfer the degree of reference makeup simply by interpolating the combination coefficients of the styles ($\alpha \in [0, 1]$) to achieve a controlled generation of makeup. We can not only gradually change the makeup style from a non-makeup image to a reference makeup image. It is also possible to interpolate two reference makeup images so that the makeup style of the non-makeup image changes from one reference makeup style to another. The concrete result is shown in FigA2.

Figure A2 The first two rows are the interpolation results of one reference makeup image, and the last two rows are the interpolation results of two reference makeup images.

A.4 Failure cases

Our network currently has some limitations, for very few reference makeup images, our network can not generate the same clear pattern. For some images with large differences in facial poses, there is a problem that the pattern on one side of the face is severely distorted, so that the pattern and the reference makeup cannot be consistent. Failure cases are shown in FigA3.

Figure A3 Failure cases.

References

1. Li T, Qian R, Dong C, et al. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In: Proceedings of the 26th ACM international conference on Multimedia. ; 2018: 645–653.

2. Jiang W, Liu S, Gao C, et al. Psgan: Pose and expression robust spatial-aware gan for customizable makeup transfer. In: ; 2020: 5194–5202.

3. Chang H, Lu J, Yu F, Finkelstein A. Pairedcyclegan: Asymmetric style transfer for applying and removing makeup. In: Proceedings of the IEEE conference on computer vision and pattern recognition. ; 2018: 40–48.

4. Liu S, Jiang W, Gao C, et al. Psgan++: Robust detail-preserving makeup transfer and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2021; 44(11): 8538–8551.

5. Wan Z, Chen H, An J, Jiang W, Yao C, Luo J. Facial attribute transformers for precise and robust makeup transfer. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. ; 2022: 1717–1726.

6. Gu Q, Wang G, Chiu MT, Tai YW, Tang CK. Ladn: Local adversarial disentangling network for facial makeup and de-makeup. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. ; 2019: 10481–10490.

7. Nguyen T, Tran AT, Hoai M. Lipstick ain't enough: beyond color matching for in-the-wild makeup transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. ; 2021: 13305–13314.

8. Xiang X, Liu D, Yang X, Zhu Y, Shen X, Allebach JP. Adversarial Open Domain Adaptation for Sketch-to-Photo Synthesis. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. ; 2022: 1434–1444.

9. Lowe DG. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 2004; 60: 91–110.

10. Tola E, Lepetit V, Fua P. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE transactions on pattern analysis and machine intelligence* 2009; 32(5): 815–830.

11. Long JL, Zhang N, Darrell T. Do convnets learn correspondence?. *Advances in neural information processing systems* 2014; 27.

12. Zhou B, Zhao H, Puig X, Fidler S, Barriuso A, Torralba A. Scene parsing through ade20k dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition. ; 2017: 633–641.

13. Choy CB, Gwak J, Savarese S, Chandraker M. Universal correspondence network. *Advances in neural information processing systems* 2016; 29.

14. Ham B, Cho M, Schmid C, Ponce J. Proposal flow: Semantic correspondences from object proposals. *IEEE transactions on pattern analysis and machine intelligence* 2017; 40(7): 1711–1725.

15. Han K, Rezende RS, Ham B, et al. Scnet: Learning semantic correspondence. In: Proceedings of the IEEE international conference on computer vision. ; 2017: 1831–1840.

16. Kim S, Min D, Ham B, Jeon S, Lin S, Sohn K. Fcss: Fully convolutional self-similarity for dense semantic correspondence. In: Proceedings of the IEEE conference on computer vision and pattern recognition. ; 2017: 6560–6569.

17. Lee J, Kim D, Ponce J, Ham B. Sfnet: Learning object-aware semantic correspondence. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. ; 2019: 2278–2287.

18. Liao J, Yao Y, Yuan L, Hua G, Kang SB. Visual attribute transfer through deep image analogy. *arXiv preprint arXiv:1705.01088* 2017.

19. Aberman K, Liao J, Shi M, Lischinski D, Chen B, Cohen-Or D. Neural best-buddies: Sparse cross-domain correspondence. *ACM Transactions on Graphics (TOG)* 2018; 37(4): 1–14.

20. Zhang P, Zhang B, Chen D, Yuan L, Wen F. Cross-domain correspondence learning for exemplar-based image translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. ; 2020: 5143–5153.

21. He M, Chen D, Liao J, Sander PV, Yuan L. Deep exemplar-based colorization. *ACM Transactions on Graphics (TOG)* 2018; 37(4): 1–16.

22. Lee J, Kim E, Lee Y, Kim D, Chang J, Choo J. Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. ; 2020: 5801–5810.

23. Zhang H, Chen W, He H, Jin Y. Disentangled makeup transfer with generative adversarial network. *arXiv preprint arXiv:1907.01144* 2019.

24. Sun Z, Chen Y, Xiong S. Ssat: A symmetric semantic-aware transformer network for makeup transfer and removal. In: . 36. Proceedings of the AAAI Conference on Artificial Intelligence. ; 2022: 2325–2334.

25. Feng Y, Wu F, Shao X, Wang Y, Zhou X. Joint 3d face reconstruction and dense alignment with position map regression network. In: Proceedings of the European conference on computer vision (ECCV). ; 2018: 534–551.

26. Park T, Liu MY, Wang TC, Zhu JY. Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. ; 2019: 2337–2346.

27. Sarfraz MS, Seibold C, Khalid H, Stiefelhagen R. Content and colour distillation for learning image translations with the spatial profile loss. *arXiv preprint arXiv:1908.00274* 2019.

28. Mao X, Li Q, Xie H, Lau RY, Wang Z, Paul Smolley S. Least squares generative adversarial networks. In: Proceedings of the IEEE international conference on computer vision. ; 2017: 2794–2802.

29. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems* 2017; 30.

30. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 2004; 13(4): 600–612.

31. Chen HJ, Hui KM, Wang SY, Tsao LW, Shuai HH, Cheng WH. Beautyglow: On-demand makeup transfer framework with reversible generative network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. ; 2019: 10042–10050.

32. Kingma DP, Dhariwal P. Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems* 2018; 31.

33. Deng H, Han C, Cai H, Han G, He S. Spatially-invariant style-codes controlled makeup transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. ; 2021: 6549–6557.

34. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. ; 2019: 4401–4410.

AUTHOR BIOGRAPHY

Mengying GAO is currently a postgraduate student with School of Computer Science, Dalian Minzu University, Dalian, China. Her research interests include computer vision.

Pengjie Wang is currently a professor with School of Computer Science, Dalian Minzu University, Dalian, China. His research interests include computer vision, graphics and data compression.