Mathematical Biosciences and Engineering

MBE, 20(1): 737−757.
DOI: 10.3934/mbe.2023034
Received: 08 August 2022
Revised: 01 October 2022
Accepted: 05 October 2022
Published: 14 October 2022

http://www.aimspress.com/journal/MBE

Research article

Large-pose facial makeup transfer based on generative adversarial network combined face alignment and face parsing

Qiming Li* and Tongyue Tu

Department of Computer Science and Technology, Shanghai Maritime University, Shanghai 201306, China

* **Correspondence:** Email: qmli@shmtu.edu.cn; Tel: +8602138282823.

Abstract: Facial makeup transfer is a special form of image style transfer. For the reference makeup image with large-pose, improving the quality of the image generated after makeup transfer is still a challenging problem worthy of discussion. In this paper, a large-pose makeup transfer algorithm based on generative adversarial network (GAN) is proposed. First, a face alignment module (FAM) is introduced to locate the key points, such as the eyes, mouth and skin. Secondly, a face parsing module (FPM) and face parsing losses are designed to analyze the source image and extract the face features. Then, the makeup style code is extracted from the reference image and the makeup transfer is completed through integrating facial features and makeup style code. Finally, a large-pose makeup transfer (LPMT) dataset is collected and constructed. Experiments are carried out on the traditional makeup transfer (MT) dataset and the new LPMT dataset. The results show that the image quality generated by the proposed method is better than that of the latest method for large-pose makeup transfer.

Keywords: makeup transfer; generative adversarial network; large-pose face; face parsing; face alignment

1. Introduction

Humanity's pursuit of beauty has never changed, and facial appearance is a crucial part of the beauty process. Among facial beautification techniques, makeup is the most popular method, along with a range of commercial products including eye shadows, lipsticks, and foundations. The desire for

beautification is rooted in social etiquette, driving the relationship between makeup and social activities in many cultures around the globe. With the development of the internet in recent years, the online sales channels of cosmetics in the world have developed rapidly. This has led to the propagation of beautiful makeup effect photos of models on the sales page of a cosmetics product. Additionally, more and more beauty makeup bloggers and live-streaming performers have begun to share their makeup appearances on various online platforms. As the online beauty and makeup market expands, consumers will be exposed to more options and will require technology to ensure beauty protocols were followed properly. Makeup transfer technology ensures these beauty protocols were followed, and functions by transferring the makeup of a makeup face photo to any plain face photo while keeping the latter's face identity information unchanged. In fact, makeup transfer technology has been widely used in many industries and fields, such as the art industry. Makeup represents people's attitude towards life and pursuit of life quality in that era. In the works left by artists of different times, the make-up of characters can help us understand the customs of that period. For the film and television industry, proper makeup can help shape the character of actors, and futuristic makeup in science fiction movies can stimulate people's imagination and exploration of future concepts. Therefore, makeup transfer is a very meaningful research direction.

Makeup transfer can be regarded as a special kind of image style transfer. Image style transfer refers to the technique of using an algorithm to learn the style of an image and then applying this style to another image. Image style transfer only involves the transfer at the domain-level, emphasizing the differences within the domain, while the makeup transfer is not only a global style transfer, but also several independent style transfers of different facial regions. Therefore, the makeup transfer requires additional makeup losses to achieve makeup details.

In recent years, as the most important and effective means and methods, machine learning and deep learning have been more and more closely combined with image processing technology. Many famous algorithms, and models of this kind, have been proposed and applied in the field of image processing, and many excellent results have been achieved. Makeup transfer has been studied for more than a decade. Early traditional makeup transfer algorithms [1−5] can be divided into two categories according to the different requirements for the data set. The first category requires a large number of paired images before and after makeup as a training set, that is, a supervised model; the second category does not require paired images, that is, an unsupervised model. The makeup transfer algorithm based on deep learning has a more straightforward framework and is the current mainstream research idea. Currently, it primarily includes the algorithms based on database matching and the algorithms based on the generative model of generate adversarial network (GAN).

The main inspiration of GAN comes from the idea of zero-sum game in game theory. When applied to deep learning neural network, it is through the continuous game between the generator network G (Generator) and the discriminant network D (Discriminator), so that G learns the distribution of data. Before GAN is used for makeup transfer, the makeup transfer method is divided into several parts and then integrated, so the output image will appear unnatural. Due to the maturity of technologies such as GAN [6−10], the facial makeup transfer algorithm has made significant progress. Most GAN-based makeup transfer algorithms use CycleGAN [11] as the primary network frame, and its structure includes two generators, G and F, and two discriminators, D and H. The image in the X domain generates the image in the Y domain through the generator G and then reconstructs back to the original image input in the X domain through the generator F; the image in the Y domain generates the X domain image through the generator F and then reconstructs it back

through the generator G. The discriminators D and H play a discriminative role in ensuring the style transfer of the images. PairedCycleGAN [12] further introduces asymmetric functions based on CycleGAN to complete the makeup transfer task. BeautyGlow [13] uses the glowing framework to divide facial information features into makeup and non-makeup features. The GAN-based makeup transfer algorithm represented by BeautyGAN [14] inputs two face images into the network, one without makeup and one with makeup. The model outputs the result after a makeup exchange, a makeup image and a makeup removal image. However, there are still some difficulties in practical applications in uncontrolled environments, such as the problem of large poses. T. Nguyen et al. [15] introduce UV texture mapping technology into the GAN framework for extremely wild makeup to improve the makeup quality of the generated images, and collect and organize the CPM dataset (Color-&-Pattern Makeup Datasets). The method proposed by Z. Sun et al. [16] adds semantic segmentation loss to the traditional makeup transfer loss, which improves the makeup effect after transfer. In addition, Z. Huang et al. [17] propose a new real-world-based automatic face makeup network IPM-Net. Z. Wan et al. [18] propose a novel Facial Attribute Transformer (FAT) and its variant Spatial FAT for high-quality makeup transfer. J. Lee et al. [19] propose to utilize the identical image with geometric distortion as a virtual reference, which makes it possible to secure the ground truth for a colored output image. However, none of the above methods has yet involved the transfer of large-pose makeup.

Large-pose makeup was proposed by W. Jiang et al. [20] in the PSGAN paper, and the makeup images with non-front perspective pose and non-neutral different expressions are collectively referred to as large-pose makeup images. Before PSGAN, makeup transfer algorithms represented by BeautyGAN [14] demonstrated success in makeup transfer. However, these methods are limited by the input condition that the source and reference images must be well-aligned front perspective faces. Performing cycle consistency [11] does not guarantee correct spatial transformations. The proposal of PSGAN [20] solves these problems. PSGAN proposes an attentive makeup morphing (AMM) module [21], which introduces an attention mechanism into the network and solves the problem of large-pose in makeup transfer, but it is easy to produce shadows in the makeup of the final transfer image. SCGAN [22] designed a flexible and controllable makeup transfer model and designed a latent space for makeup extraction of large poses. However, this method is simple to process the source image, and results in several issues including missing facial features and blurred and inaccurate makeup in the image after the final makeup transfer.

In this paper, by adopting the makeup style extraction module and makeup integration module in SCGAN, a large-pose makeup transfer model is proposed. First, a face alignment module (FAM) is designed and introduced to locate the key points of face better. Secondly, a face parsing module (FPM) is designed for the original facial feature extraction encoder, which introduces a convolutional neural network to improve the accuracy of facial identity feature extraction, and proposes a face parsing loss. The FPM includes two branches, the face feature extraction branch and the face reconstruction branch. The former is used to extract face features, and the latter is used to reconstruct a face based on the extracted features and generate an image. Face parsing loss is used to constrain the reconstructed face to be similar to the input face. The two branches work together to improve the accuracy of face feature extraction and maintain the consistency of face identity. In addition, since there are few data sets based on large-pose makeup images, in this paper, a new makeup transfer data set mainly composed of large-pose makeup images (abbreviated as LPMT) is collected and established for experiments.

2. Materials and methods

2.1. Dataset

Two datasets are used in this paper: the makeup transfer (MT) dataset [14] and the LPMT dataset. The MT dataset is published by BeautyGAN [14], with a total of 3834 face images, including 2719 makeup images and 1115 non-makeup images. Since most of the makeup images in the MT dataset are based on normal-pose images, to conduct the experiment, 100 makeup images with different poses and expressions are selected manually. PSGAN [20] has collected and organized a new makeup transfer dataset for large poses and multiple expressions, namely the makeup-wild dataset, which contains 403 makeup images and 369 non-makeup images. However, the number of images in the dataset is small, and the dataset cannot be found on PSGAN's official website. The facial poses and expressions of the images in the LPMT dataset are very diverse. To collect this dataset, we first searched for a set of makeup images using keywords (e.g., Japanese makeup, European and American makeup, classical makeup, sweet makeup, etc.), selected makeup images with non-frontal and non-neutral expressions, and then uniformly converted the collected images into .png format and crop out the face area in each image. Finally, we manually removed poor quality and inappropriate face images and obtained 2854 makeup images and 343 non-makeup images.

Figure 1. Large-pose: (1) makeup; (2) non-makeup.

Figure 2. Normal-pose: (1) makeup; (2) non-makeup.

In this paper, the quality of makeup images generated will be analyzed. So, the face regions should be cropped out in a rectangular pattern from all images. Some images of large-pose makeup and no makeup are shown in Figure 1, and some images of normal-pose makeup and no makeup are shown in Figure 2.

2.2. Methods

2.2.1. Formulation

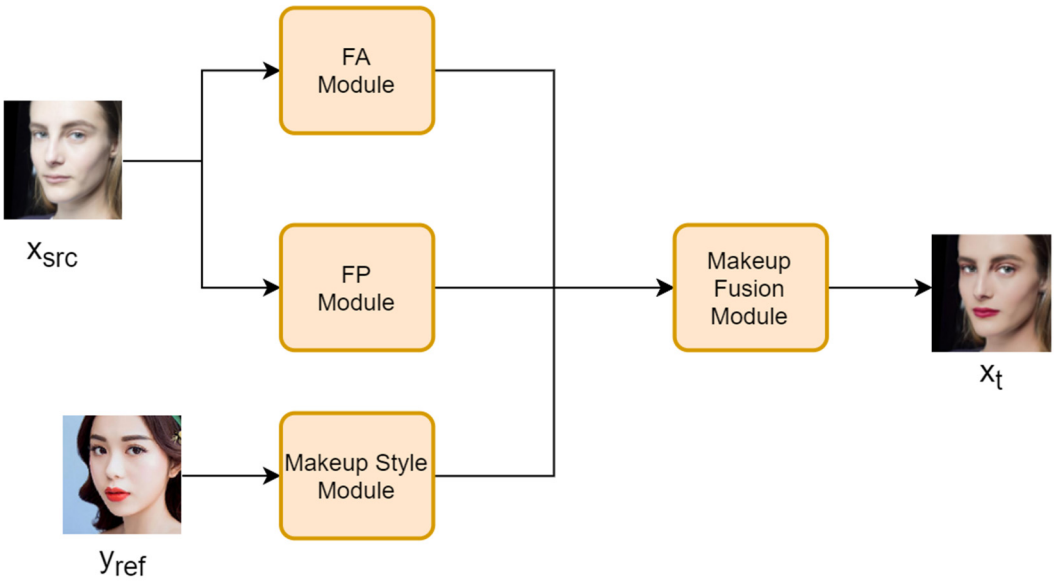

Figure 3. Network structure.

Let X represents the source image domain and Y represents the reference image domain. Given any source image $x_{src} \in$ X and any reference image $y_{ref} \in$ Y, the goal of the algorithm is to transfer the makeup style from the reference image y_{ref} to a source image x_{src} and get the target image x_t, so that x_t has the makeup style of y_{ref} and the face information features of x_{src}. As shown in Figure 3, the overall network structure is divided into four parts: the FAM, FPM, makeup style module and makeup fusion module. The makeup style module [22] extracts the makeup information of the reference image. It maps it to a one-dimensional style code Z. The FAM locates the key parts of the face. The FPM extracts the face identity features of the source image. The makeup fusion module [22] fuses makeup style and facial identity features and generates final results.

2.2.2. Face alignment module

Face alignment is used to locate the facial key point features. Before face parsing, face preprocessing is very important. We need to detect the face in the image, and then obtain the aligned standard face through face similarity transformation. At present, the most commonly used alignment method is to detect five face key points through CNN, and then use similarity transformation to obtain the aligned face. The FAM introduced to our method is used to accurately locate the face features, improve the fusion accuracy of the face and the makeup code, and improve the final makeup effect.

The face alignment module consists of three convolutional layers, three max-pooling layers and one fully connected layer [23], and the structure is shown in Figure 4.

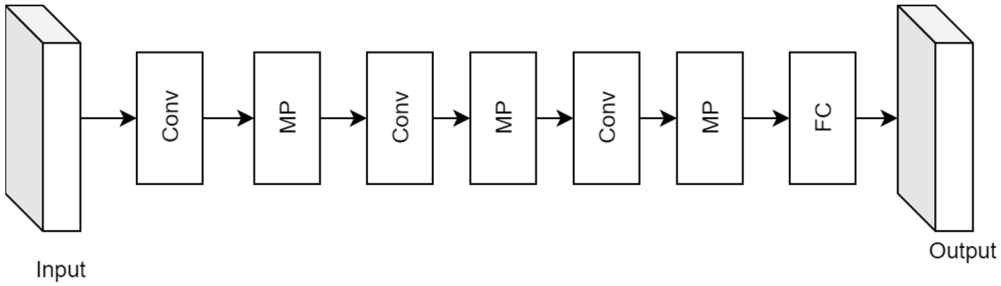

Figure 4. Structure of face alignment module.

2.2.3. Face parsing module

With the development of neural networks, models based on convolutional neural networks have increased rapidly [24−26]. In this paper, the FPM is proposed to extract face identity feature information accurately from the source image. The accuracy of facial feature extraction will directly affect the quality of the final generated image. FPM extracts the facial features of input image x, which is recorded as:

$$F_{id} = FPM(x). \tag{1}$$

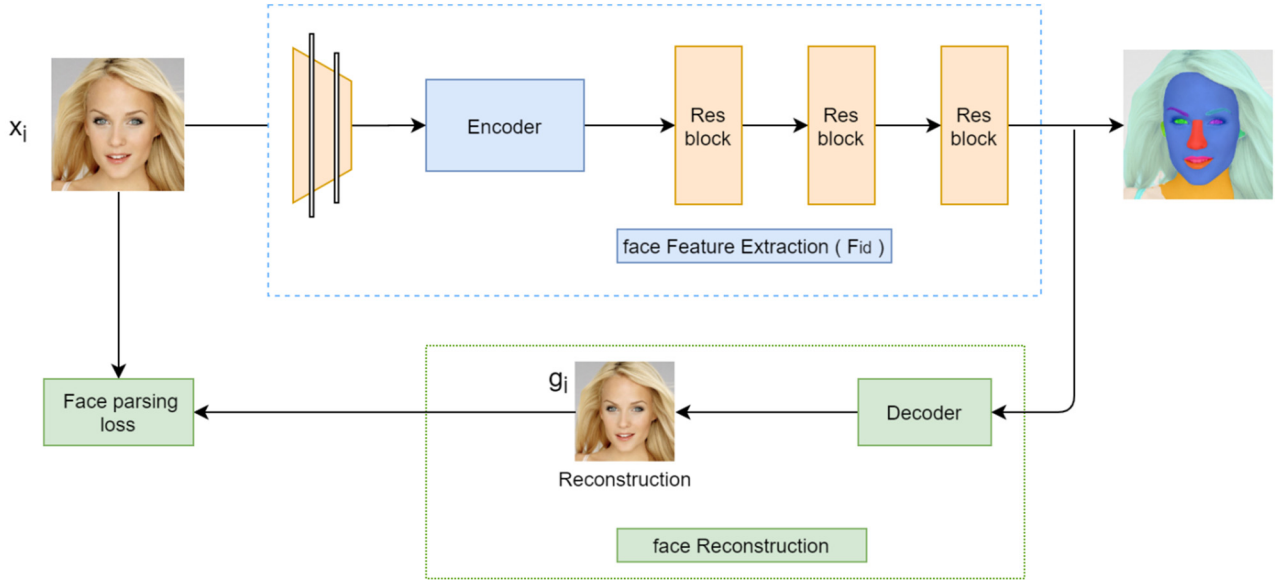

Figure 5. Structure of face parsing module.

As shown in Figure 5, the FPM consists of two branches: the face parsing branch and the face reconstruction branch. The former consists of two down-sampling layers, an encoder and three residual blocks [27]. The encoder consists of three convolutional modules. The structure of the convolution module [28] is shown in Figure 6. Each convolution module contains a 1 × 1 convolution layer, a Batch Normalization layer, a ReLU activation layer and a max-pooling layer. The parsing branch improves

the accuracy of facial feature extraction by introducing convolutional neural networks. The face reconstruction branch contains a decoder consisting of deconvolution layers and decodes the face image based on the extracted face features. In this paper, the reconstruction loss is used to constrain the generated face, which makes the face identity information of the reconstructed generated image consistent with that of the input source image, and further improves the accuracy of face feature extraction. The two branches need to be run in the training phase, and only the face parsing branch needs to be run in the calculation phase.

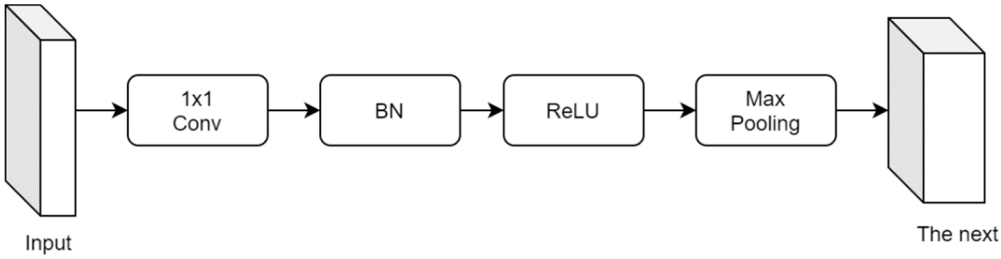

Figure 6. Convolution module.

2.2.4. Makeup style module

The structure of makeup style module (MSM) is shown in Figure 7, which consists of two down-sampling layers, a mapping module [29], and a latent space [22]. Style codes can be obtained by simply averaging facial features into a one-dimensional vector, but such style codes follow similar probability densities of training data, leading to inevitable entanglement between facial components; a nonlinear mapping module can solve the problem. First, the reference makeup image y_{ref} is decomposed into three parts: y_{lip}, y_{eyes} and y_{skin} through the face parser. Second, the initial style codes are embedded in the latent space by the nonlinear mapping module, so the generated style codes are not restricted by the distribution of training data and can be decomposed.

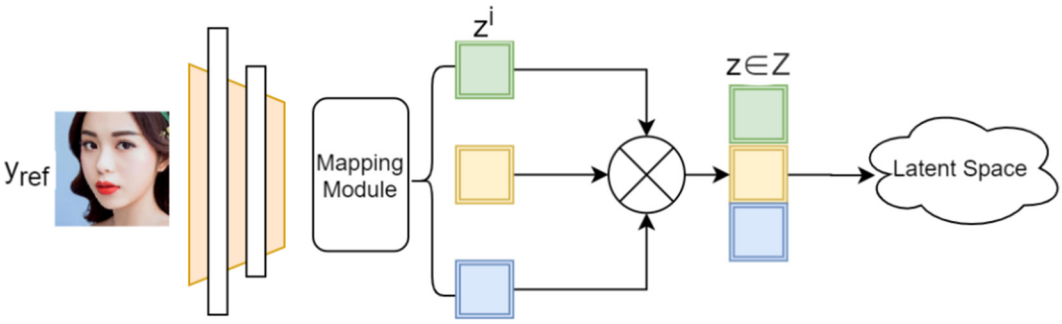

Figure 7. Structure of makeup style module.

The specific structure of the mapping module is shown in Figure 8, which includes an average pooling layer and a 1 × 1 convolutional layer. Each component y_i is mapped to the style code z_i through the mapping module to obtain z_{lip}, z_{eyes} and z_{skin}. The style codes of the three components are then concatenated to get the complete initial style code Z in the latent space. The makeup code Z in the

potential space only retains the color information of the makeup, ignoring other irrelevant information (such as mouth opening or closing, eyes widening or narrowing, etc.) to realize the makeup transfer of large-pose.

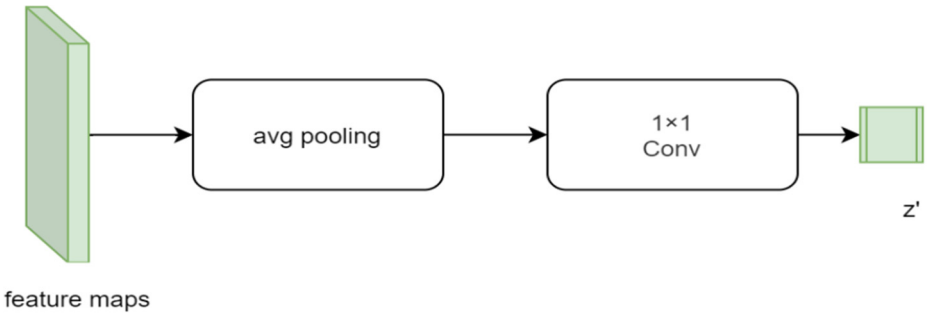

Figure 8. Mapping module.

2.2.5. Makeup fusion module

The structure of makeup fusion module (MFM) is shown in Figure 9(a), which includes two down-sampling layers and three fusion blocks. Its function is to combine the makeup style code Z and the facial features F_{id} to generate a target image with the makeup style of the reference image and the facial feature information of the source image. The structure of the fusion block [22] is shown in Figure 9(b), including two convolutional layers, two AdaIN layers, and a ReLU layer, where each AdaIN [30] layer is used to connect Z and F_{id}, and finally get the target transfer image.

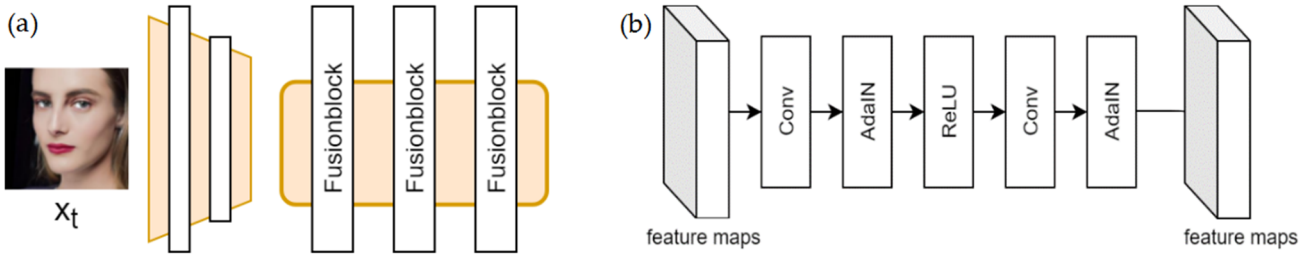

Figure 9. (a) Structure of makeup fusion module; (b) Structure of fusion block.

2.3. Loss function

Since there is no pairing between makeup and non-makeup images, a circular approach is used to train the generative adversarial network. Given a non-makeup image x and a makeup image y, the generative adversarial network is used to realize the mutual mapping between X domain and Y domain. Hence, the method of CycleGAN [12] is adopted in the training process, and the loss function includes adversarial loss, cycle consistency loss, perceptual loss, makeup loss, and face parsing loss.

2.3.1. Adversarial loss

The most basic loss of GAN [31] is used to guide the generator to generate more realistic images. In this paper, two discriminators, D_X and D_Y, are used, and the output value is [0,1], where $D_X = 1$ indicates that the output image is from the *X* space, and $D_Y = 1$ suggests that the output image is from the *Y* space. The purpose of GAN is to make discriminator learn the distribution of the data. The two discriminators both adopt the structure of PatchGAN [11]. The adversarial loss of the generator $\mathcal{L}_{\mathrm{G}}^{\mathrm{adv}}$ and the adversarial loss of the discriminator $\mathcal{L}_{\mathrm{D}}^{\mathrm{adv}}$ are respectively defined as:

$$\mathcal{L}_D^{adv} = -E_{x\sim X}[logD_X(x)] - E_{y\sim Y}[logD_Y(y)] - E_{x\sim X,y\sim Y}\left[\log\left(1 - D_X\left(G(y,x)\right)\right)\right] - E_{x\sim X,y\sim Y}\left[\log\left(1 - D_Y\left(G(x,y)\right)\right)\right], \quad (2)$$

$$\mathcal{L}_G^{adv} = -E_{x\sim X,y\sim Y}\left[\log\left(D_X\left(G(y,x)\right)\right)\right] - E_{x\sim X,y\sim Y}\left[\log\left(D_Y\left(G(x,y)\right)\right)\right], \quad (3)$$

where $E_{x\sim X}$ means taking samples from *X* space, $E_{y\sim Y}$ means taking samples from *Y* space, *E*(.) represents the expected value of the distribution function, and *G*(.) represents the image generated by generator *G*.

2.3.2. Cycle consistency loss

Since the source image is not paired with the reference image, in order to make the final generated image have both the facial identity features of the source image and the makeup style of the reference image, the cycle consistency loss proposed by CycleGAN [11] is adopted. Since two generators are used, the network structure should learn the mapping of these two generators at the same time, and hope that $G(G(y,x),y)$ is as similar to y as possible, and $G(G(x,y),x)$ is as similar to x as possible, and the L1 norm is used to constrain the reconstructed image. The cycle consistency loss $\mathcal{L}_{cyc}$ is defined as:

$$\mathcal{L}_{cyc} = \| G(G(x,y),x) - x) \|_1 + \| G(G(y,x),y) - y) \|_1, \quad (4)$$

where $\|.\|_1$ represents the L1 norm, *x* is the source image, *y* is the reference image, and *G* is the generator.

2.3.3. Makeup loss

The makeup style varies from person to person, and makeup transfer is more about the transformation of independent styles in different areas of the face, so makeup loss is introduced. The makeup loss is proposed by BeautyGAN [14], which is calculated by Histogram Matching (HM) to improve the makeup details of the generated images. The histogram matching loss is more stable than the Gram loss used in the style layer of traditional style transfer. The makeup loss $\mathcal{L}_{makeup}$ is defined as:

$$\mathcal{L}_{makeup} = \| G(x,y) - HW(x,y) \|_2 + \| G(y,x) - HW(y,x)) \|_2, \quad (5)$$

where $\|.\|_2$ represents the L2 norm, $HW(.)$ denotes histogram matching loss, and $HW(x,y)$

denotes the makeup style with y while preserving the identity features of x.

2.3.4. Perceptual loss

Perceptual loss is used to ensure that the facial detail features of the generated image are similar to the source image and improve the quality of the generated image. The perceptual loss proposed by J. Johnson, et al. [32] is adopted, which is computed from the high-dimensional feature distance between two images. The L2 loss is used to measure their differences. The perceptual loss L_{per} is defined as:

$$\mathcal{L}_{per} = \| F_l\big(G(y,x)\big) - F_l(y) \|_2 + \| F_l\big(G(x,y)\big) - F_l(x) \|_2, \tag{6}$$

Among them, $F_l(.)$ is the output of the first layer of the VGG-16 [25] model, and the VGG-16 model adopts the pre-trained model on ImageNet.

2.3.5. Face parsing loss

To accurately extract the face identity feature information of the source image, the face parsing loss is introduced into the FPM module. The parsing loss is used to constrain the generated images to be similar to the input ones. Assuming that the input image is x, the image generated after reconstruction is g, and n is the number of pixels, The parsing loss is defined as:

$$\mathcal{L}_{fp} = \frac{1}{n}\sum_{i=1}^{n}(g_i - x_i)^2, \tag{7}$$

2.3.6. Total loss

In summary, the overall loss of the entire network structure is defined as:

$$\mathcal{L}_{total} = \lambda_{adv}\big(\mathcal{L}_D^{adv} + \mathcal{L}_G^{adv}\big) + \lambda_{cyc}\mathcal{L}_{cyc} + \lambda_{makeup}\mathcal{L}_{makeup} + \lambda_{per}\mathcal{L}_{per} + \lambda_{fp}\mathcal{L}_{fp}, \tag{8}$$

where $\lambda_{\cdot}$ is the weight of the respective loss function.

2.4. Implementation details

In this paper, combined with the operation methods of the datasets in BeautyGAN [14] and PSGAN [20], 100 makeup images are randomly selected from the MT dataset, and 150 makeup images and 100 non-makeup images are randomly selected from the LPMT dataset for testing. The rest of the LPMT dataset images are used as the training set for the experiment. In the test phase, the large-pose makeup is tested, and the normal-pose face makeup is also tested.

All experiments in this paper are trained and tested on NVIDIA GTX 1650Ti GPU through Pytorch 1.6.0. Features are extracted from the Relu_4_1 layer of VGG16 [25], and the perceptual loss is calculated. The optimizer for generator and discriminator is Adam [33], where β_1 = 0.2, β_2 = 0.9, learning efficiency is set to 0.0002, and batch size is set to 1. The weights of each loss function are set as λadv = λmakeup = 1, λcyc = 10, λper = 0.005 and λfp = 0.01. In order to ensure the quality

of the final generated images and to better compare with previous methods in the comparison experiment, the weights of the loss functions here are consistent with that of previous methods.

3. Results

3.1. Makeup and removal

In the experimental test process, simultaneously inputting a makeup image and a non-makeup image can get the image after makeup transfer and an image after makeup removal. In this paper, normal-pose makeup and large-pose makeup are tested successively, and the final results are shown in Figure 10. In both (a) and (b) of Figure 10, the first column is the image without makeup; the second column is the image with makeup, the third column is the image with makeup transferred, and the fourth column is the image with makeup removed.

Figure 10. (a) Normal-pose makeup images; (b) Large-pose makeup images.

3.2. Comparative experiment

3.2.1. Comparison of normal-pose makeup transfer

For the reference images of normal-pose makeup, the method proposed in this paper is compared with the current advanced makeup transfer methods, and the results are shown in Figures 11. Since BeautyGAN [14], BeautyGlow [13], PairedCycleGAN [12], and SOGAN [34] did not disclose the code and the trained model, the result pictures given in the references are directly used for comparison.

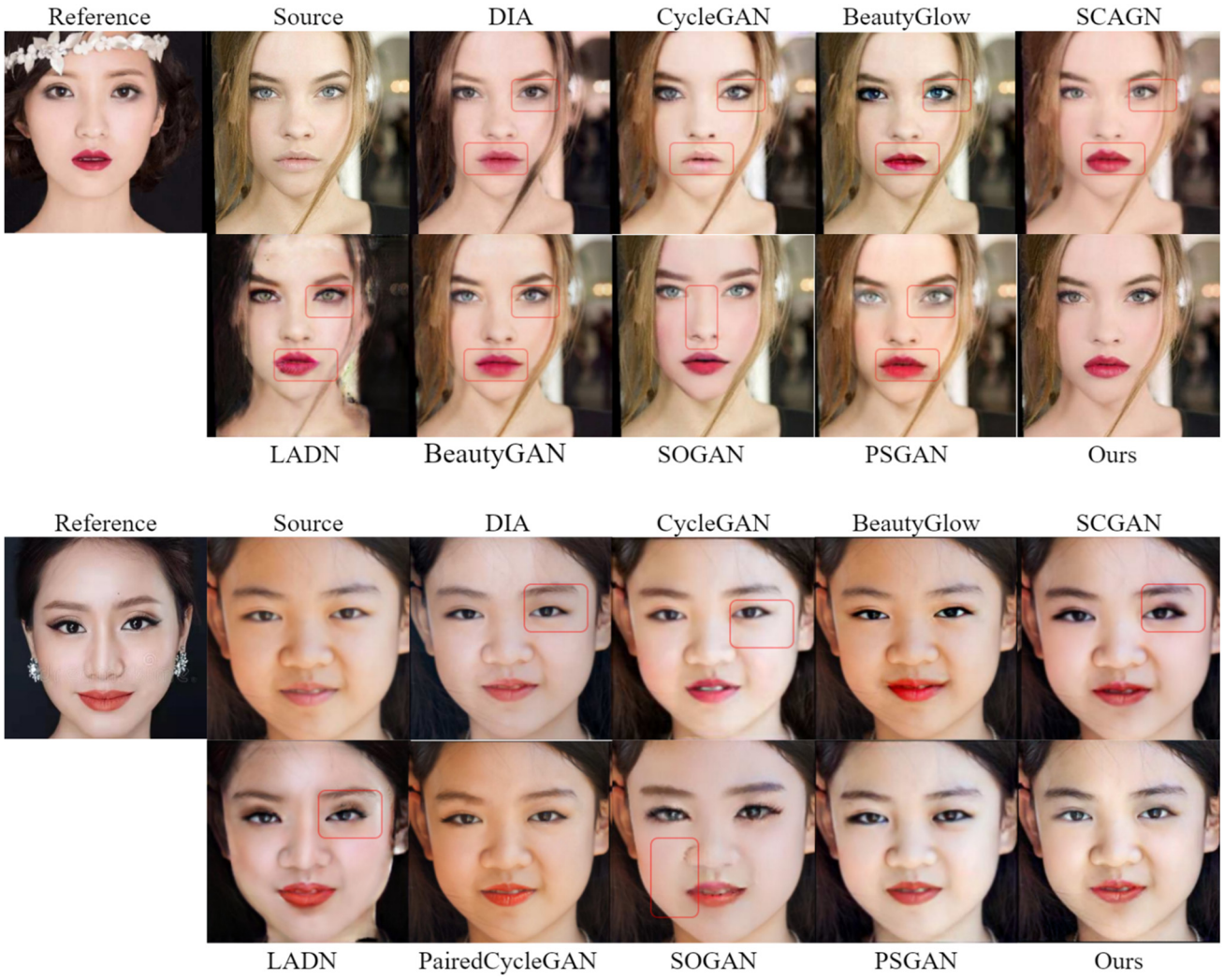

Figure 11. Comparison of normal-pose makeup transfer.

As can be seen from the two groups of comparison images, the images generated by DIA [35] have very light colors after the transfer of eye and lip makeup, and the makeup sense is not apparent enough. The images generated by the CycleGAN [11] method have a relatively weak sense of makeup, and there are no specific makeup details. Most obviously, the lip color of the generated images in the first comparison failed to transfer successfully. The images generated by PairedCycleGAN [12] method also had a weak sense of makeup, and the skin color of the images generated by this method did not change in the second group. BeautyGAN [14] and BeautyGlow [13] generate better images than the previous three methods, but there is still room for improvement. In the first set of comparison images, the lip shape of the image generated by BeautyGAN is not obvious; in the image generated by BeautyGlow, the red lip makeup has faint signs of black, and the skin color is still the paler skin color in the source image. In the second set of comparison images, the skin color of the image generated by BeautyGlow still failed to transfer successfully, and it is still the darker yellow skin color in the source image. The face identity information of the images generated by the LADN [36] method is slightly changed compared to the source images. Especially in the first set of comparison images, the image generated by the LADN method has rough makeup, and the skin color of the forehead is layered. The

images generated by SOGAN [34] have a good makeup effect and good makeup sense, but there are also some changes in facial identity information. In the first set of comparison images, the nose tip of the image generated by the SOGAN method has signs of whitening, and the lip shape also changes; in the second set of comparison images, the facial nasolabial folds of the image generated by SOGAN become lighter. The image makeup effect generated by PSGAN [20] and SCGAN [22] methods is better, but the makeup details can be improved. In the first set of comparison images, the eye makeup of the image generated by the PSGAN method is very light, and the black eyeliner is not obvious enough; the lip makeup of the image generated by the SCGAN method is rough, and the lip peak also appears red. In the second set of comparison images, the black eye shadow of the image generated by the SCGAN method is purple. In contrast, the lip makeup of the image generated by our method in the first row fits the lip shape better, and the details of the eye makeup part are also better. In the image generated by our method in the second row, the eye makeup is cleaner. Both images also well preserve the identity feature information of the source image. Because the addition of FPM and FAM can more accurately locate the facial features of the face in the image, it is not only suitable for large-pose faces, but also suitable for normal-pose faces so that when the makeup is transferred, the color of the makeup can more fit the contours of the facial features, showing better makeup details.

3.2.2. Comparison of large-pose makeup transfer

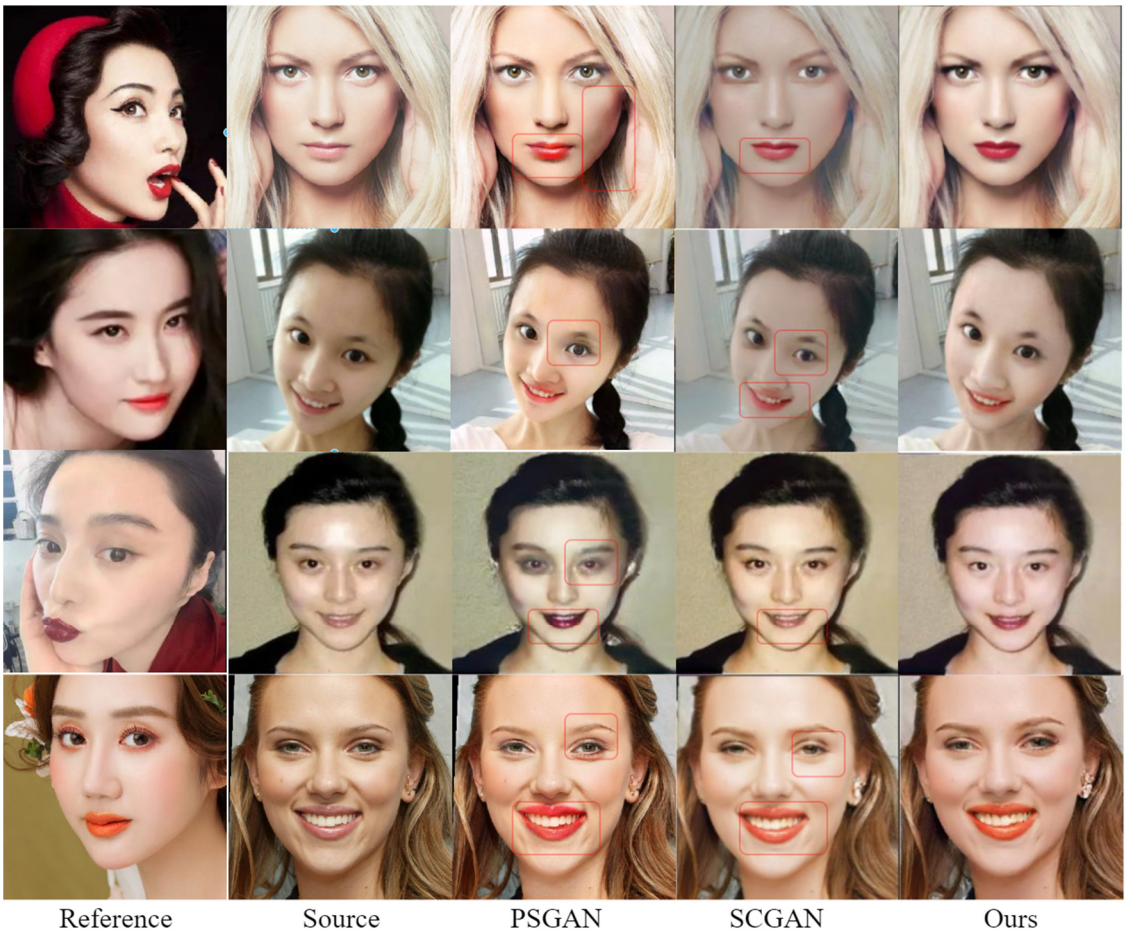

Figure 12. Comparison of large-pose makeup transfer.

For reference images of large-pose makeup with different expressions and poses, the method proposed in this paper is compared with the current state-of-the-art methods SCGAN [22] and PSGAN [20]. The results are shown in Figure 12. It can be seen from the figure that the makeup of the images generated by the PSGAN [20] method in the first row, the second row, and the third row have serious shadow marks, and the lip color of the generated image in the fourth row is wrong. The images generated by the SCGAN [22] method are prone to problems such as failure to apply some makeup, or the makeup is not detailed enough. As demonstrated by the skin color among them, the skin color of the images generated in the first and second rows is wrong, the skin color of the third row does not change, the lip color is also lighter, and the eye makeup in the fourth row is not obvious enough. Detailed in the overall comparison, through adding FPM and FAM in our method, the generated image in the first row does not show black skin color on the right face, and the black eyeliner is also more obvious. Additionally, the generated images in the second and third rows do not have the phenomenon of "dark circles". In the fourth row, the image generated by the method in this paper shows the details of the orange blush, and the match between the color of the lipstick and the lip shape is also better.

3.3. Objective and subjective evaluation

3.3.1. Objective evaluation

Most makeup transfer papers do not use objective indicators to evaluate the experimental results. In this paper, based on the experience of CPM [15] and AesGAN [37], two objective evaluation indicators, namely structural similarity index measure (SSIM) and peak signal to noise ratio (PSNR), are used. The quality is objectively assessed and tested on the MT and LPMT datasets. PSNR is the most common and widely used objective method for evaluating the quality of generated images. SSIM is often used to evaluate image quality. The higher the scores of these two objective indicators demonstrates better quality of the image. The final average score of each indicator is shown in Table 1. Since our process handles makeup details better, the average score for both metrics is the highest.

Table 1. Objective index evaluation results.

Methods	LPMT Dataset		MT Dataset	
	PSNR	SSIM	PSNR	SSIM
PSGAN	23.55db	0.732	23.83db	0.765
SCGAN	30.71db	0.802	33.45db	0.886
Ours	**32.36db**	**0.853**	**36.66db**	**0.918**

3.3.2. Subjective evaluation

In this paper, 10 participants are invited to score the generated makeup images subjectively. The 10 participants included eight girls and two boys, all undergraduates or graduate students. The eight girls are all familiar with beauty makeup, and the two boys have limited knowledge of makeup applications. When scoring and evaluating, it is required to consider the makeup effect and the unchanged face identity feature information, of which 5 is the highest score, 1 is the lowest score, and the final average is taken. The results are shown in Table 2, and the improved method in this paper is

better recognized.

Table 2. Subjective index evaluation results.

Participants	MT Dataset			LPMT Dataset		
	PSGAN	SCGAN	Our	PSGAN	SCGAN	Our
Participant 1	3	3	4	3	3	3
Participant 2	3	3	4	3	3	3
Participant 3	3	4	4	3	4	4
Participant 4	3	3	4	3	3	4
Participant 5	3	3	4	3	3	4
Participant 6	3.5	3.5	3.5	3	3.5	3.5
Participant 7	3.5	3.5	3.5	3	3.5	3.5
Participant 8	3.5	3.5	4	3.5	3.5	4
Participant 9	4	4	4	3.5	3.5	4
Participant 10	4	4	3.5	3.5	3.5	3.5
Average	3.35	3.45	**3.85**	3.15	3.35	**3.65**

3.4. Ablation experiment

The ablation experiments are carried out for the functions of the face parsing module and the face alignment module.

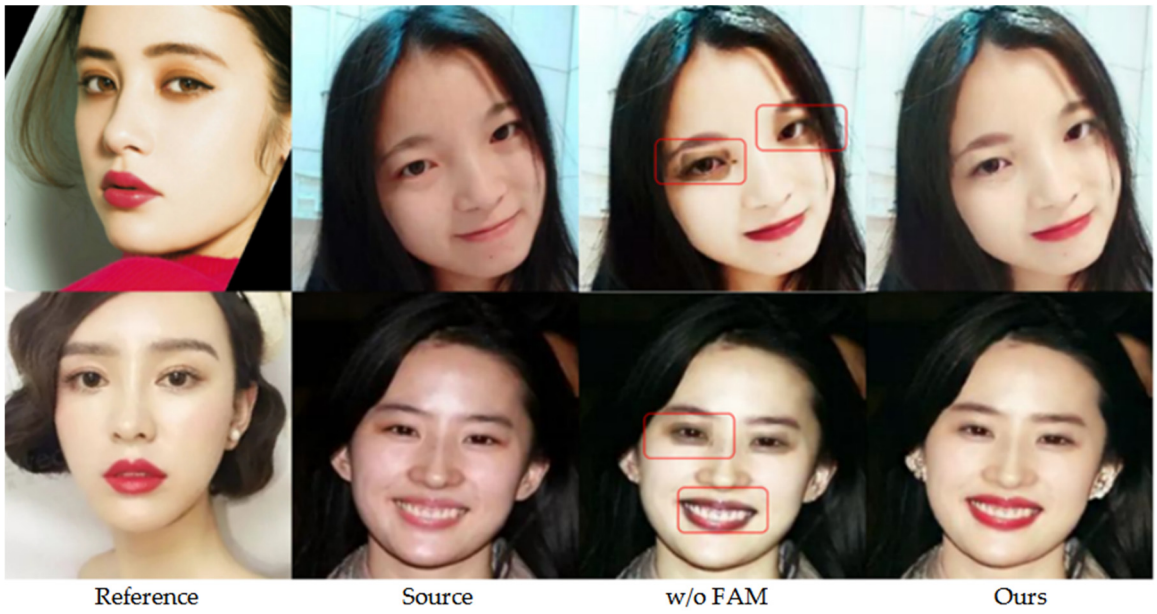

Figure 13. Ablation experiment for FAM.

When the face alignment module is not added, the final generated image will fail to apply makeup in some parts due to the inaccurate positioning of face features. As shown in Figure 13, in the first row, the image lacking FAM has the problem of missing eye makeup. In the second row, the image lacking FAM not only lacks eye makeup, but also has red lips turning black lips in the lip

makeup. This is because the facial features of large-pose face images are more difficult to accurately locate than the normal-pose face images. The SCGAN method does not perform additional face alignment processing on the source image and the reference image before makeup transfer, so the extraction of the identity feature information of the source image may fail, and the makeup of the reference image may also fail to be extracted, resulting in partial facial makeup transfer failure when the final makeup is fused. When FAM is added, the accuracy of facial feature positioning is improved, and the problem of lack of facial makeup is also solved.

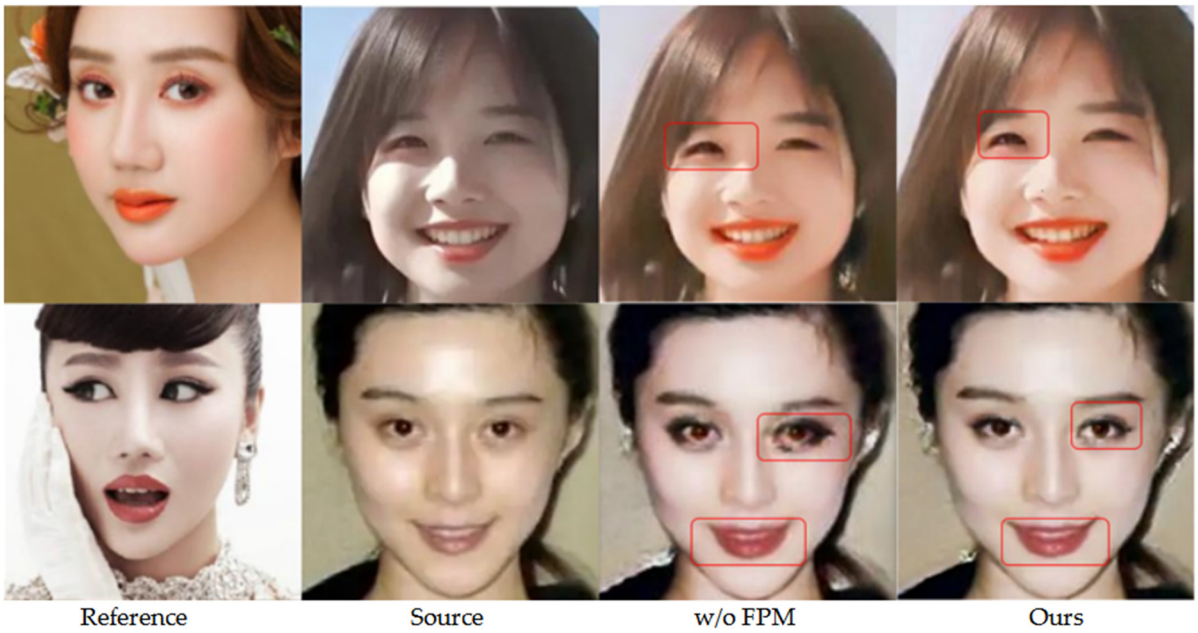

Figure 14. Ablation experiment for FPM.

Without adding the face parsing module, some parts of the face in the generated makeup images are prone to uneven coloring. As shown in Figure 14, the makeup effect after adding the face parsing module is more delicate and natural. First, the generated image in the first row and third column does not have FPM added, the eyes in the image are covered with black eyeshadow, and the area other than the lips is also stained with orange lipstick. In the first row and fourth column of the generated image with FPM, the white part of the eye can still be seen clearly at the right eye, and the color of the lipstick is also very suitable for the shape of the lips. Secondly, in the second row and third column of the generated images without FPM, the distribution of black eye shadow is uneven, and the lipstick color is not enough to fit the lip shape. In the fourth column of the second row, the final generated images of FPM are added, the eye shadow is clean and even, and the lip color is more suitable to the lip shape.

Figure 15 shows the face segmentation diagram. Pixels of different colors represent different facial components. First, in the comparison image in the first row, the extraction of the eye region of the left eye in the face segmentation without FPM is too large and not precise enough, and the final eye makeup effect is not as good as the makeup effect after adding FPM. Secondly, in the face segmentation image without FPM in the second line, the extraction of the lip region is not accurate, and the final makeup effect of the lip is not as good as that after FPM is added.

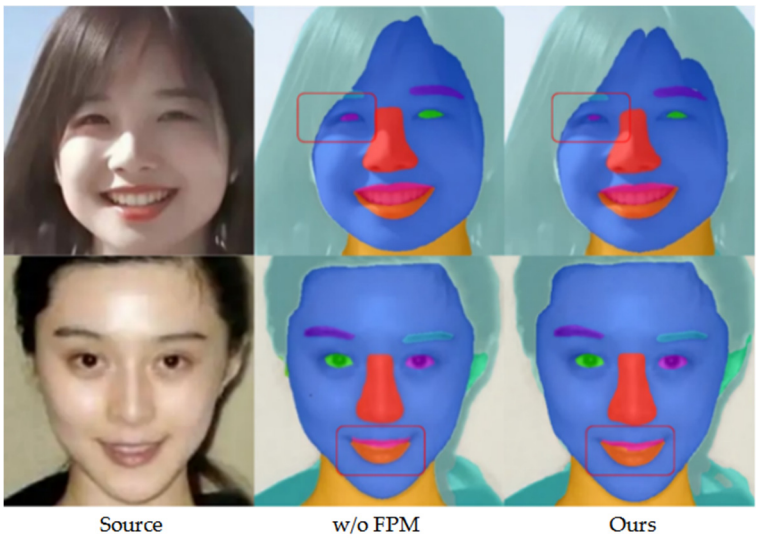

Figure 15. Face segmentation for FPM.

4. Discussion

There are three conclusions that can be reached after analyzing our experimental results. First, the main purpose of this work is to improve the makeup transfer of large-pose faces. However, there are few open-source large-pose facial makeup transfer datasets available at present. Therefore, LPMT datasets are collected and built. Subsequently, the comparative experiments with previous classical algorithms are conducted on both LPMT dataset and the normal-pose facial makeup transfer dataset, MT. This makes our results more convincing. Second, our improved algorithm has achieved better results in both normal-pose and large-pose makeup transfer. Makeup transfer pays more attention to makeup details, and the improvement of makeup details can make the overall makeup effect better. The images generated by our method are more delicate in makeup details and solve the makeup transfer failure problem in some images. Lastly, most makeup transfer algorithms do not take objective indicators to evaluate the generated images, SSIM and PSNR are used in this paper to evaluate the experimental results and achieve good results.

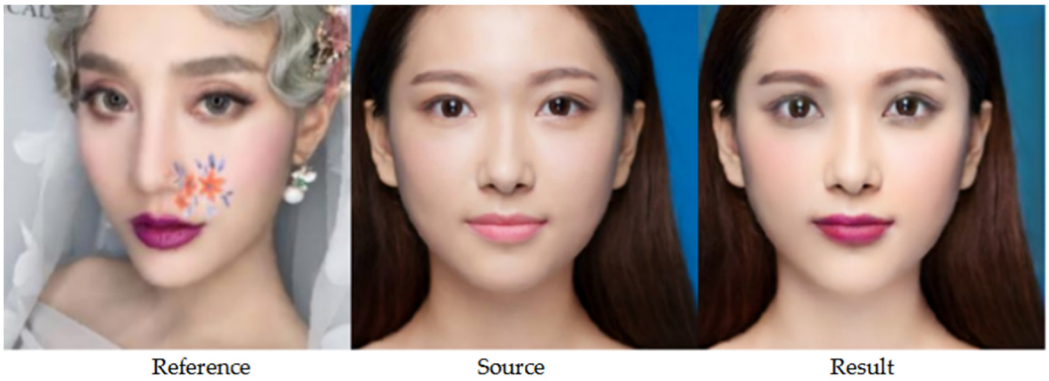

Figure 16. Limitation.

Although our improved method improves the effect of makeup transfer, it still has the disadvantage that it cannot transfer the pattern part on the face makeup. As shown in Figure. 16, our method failed to move the face pattern in the reference image to the final generated image. Facial patterns include stickers, tattoos, jewelry, etc., but the current makeup transfer framework only focuses on the operation of facial makeup color, and does not involve the extraction and transfer of these facial patterns. In order to better transfer makeup for large-pose makeup, our model only retains color information when extracting makeup, ignoring other information such as shape, and cannot extract pattern features. The transfer of facial patterns may be applied to 3D face model technology, which will be one of our future improvement research directions.

5. Conclusions

In this paper, a makeup transfer algorithm based on a generative adversarial network is proposed by introducing and combining the face alignment module and the face parsing module. In the face alignment module, the fusion accuracy of the face and makeup code is improved by accurately locating the face features. In the face parsing module, a convolutional neural network and a face reconstruction branch are introduced to improve the accuracy of face feature extraction. The makeup style module extracts makeup codes from the reference images, and the makeup fusion module fuses makeup codes with facial identity features to complete makeup transfer. These modules solve problems together and achieve a good effect of makeup transfer. In addition, a new dataset of large-pose makeup is proposed for experimental study.

Acknowledgments

This research was funded by the National Natural Science Foundation of China, grant number 41701523 and Natural Science Foundation of Shanghai, China, grant number 14ZR1419700.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. W. S. Tong, C. K. Tang, M. S. Brown, Y. Q. Xu, Example-based cosmetic transfer, in *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, (2007), 211−218. https://doi.org/10.1109/PG.2007.31
2. S. Liu, X. Ou, R. Qian, W. Wang, X. Cao, Makeup like a superstar: deep localized makeup transfer network, preprint, arXiv:1604.07102.
3. D. Guo, T. Sim, Digital face makeup by example, in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (2009), 73−79. https://doi.org/10.1109/CVPR.2009.5206833
4. L. Xu, Y. Du, Y. Zhang, An automatic framework for example-based virtual makeup, in *2013 IEEE International Conference on Image Processing*, (2013), 3206−3210. https://doi.org/10.1109/ICIP.2013.6738660

5. C. Li, K. Zhou, S. Lin, Simulating makeup through physics-based manipulation of intrinsic image layers, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2015), 4621−4629. https://doi.org/10.1109/CVPR.2015.7299093
6. W. Xu, C. Long, R. Wang, G. Wang, DRB-GAN: A dynamic resblock generative adversarial network for artistic style transfer, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, (2021), 6383−6392. https://doi.org/10.1109/ICCV48922.2021.00632
7. H. Chen, L. Zhao, H. Zhang, Z. Wang, Z. Zuo, A. Li, et al., Diverse image style transfer via invertible cross-space mapping, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, (2021), 14860−14869. https://doi.org/10.1109/ICCV48922.2021.01461
8. Y. Hou, L. Zheng, Visualizing adapted knowledge in domain transfer, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 13824−13833. https://doi.org/10.1109/CVPR46437.2021.01361
9. X. Zhang, Z. Cheng, X. Zhang, H. Liu, Posterior promoted GAN with distribution discriminator for unsupervised image synthesis, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 6519−6528. https://doi.org/10.1109/CVPR46437.2021.00645
10. P. Wang, Y. Li, N. Vasconcelos, Rethinking and improving the robustness of image style transfer, in 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 124−133. https://doi.org/10.1109/CVPR46437.2021.00019
11. J. Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in *2017 IEEE International Conference on Computer Vision (ICCV)*, (2017), 2223−2232. https://doi.org/10.1109/ICCV.2017.244
12. H. Chang, J. Lu, F. Yu, A. Finkelstein, Pairedcyclegan: Asymmetric style transfer for applying and removing makeup, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 40−48. https://doi.org/10.1109/CVPR.2018.00012
13. H. J. Chen, K. M. Hui, S. Y. Wang, L. W. Tsao, H. H. Shuai, W. H. Cheng, Beautyglow: On-demand makeup transfer framework with reversible generative network, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019), 10042−10050. https://doi.org/10.1109/CVPR.2019.01028
14. T. Li, R. Qian, C. Dong, S. Liu, Q. Yan, W. Zhu, 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. https://doi.org/10.1145/3240508.3240618
15. T. Nguyen, A. T. Tran, M. Hoai, Lipstick ain't enough: beyond color matching for in-the-wild makeup transfer, in 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 13305−13314. https://doi.org/10.1109/cvpr46437.2021.01310
16. Z. Sun, Y. Chen, S. Xiong, SSAT: A symmetric semantic-aware transformer network for makeup transfer and removal, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **36** (2022), 2325−2334. https://doi.org/10.1609/aaai.v36i2.20131
17. Z. Huang, Z. Zheng, C. Yan, H. Xie, Y. Sun, J. Wang, et al., Real-world automatic makeup via identity preservation makeup net, in *International Joint Conferences on Artificial Intelligence Organization*, (2020), 652−658. https://doi.org/10.24963/ijcai.2020/91

18. Z. Wan, H. Chen, J. An, W. Jiang, C. Yao, J. Luo, et al., Facial attribute transformers for precise and robust makeup transfer, in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, (2022), 1717−1726. https://doi.org/10.1109/wacv51458.2022.00317
19. J. Lee, E. Kim, Y. Lee, D. Kim, J. Chang, J. Choo, Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), 5800–5809. https://doi.org/10.1109/cvpr42600.2020.00584
20. W. Jiang, S. Liu, C. Gao, J. Cao, R. He, J. Feng, et al., Psgan: Pose and expression robust spatial-aware gan for customizable makeup transfer, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), 5194–5202. https://doi.org/10.1109/CVPR42600.2020.00524
21. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, in *Advances in Neural Information Processing Systems*, **30** (2017), 5998−6008.
22. H. Deng, C. Han, H. Cai, G. Han, S. He, Spatially-invariant style-codes controlled makeup transfer, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 6549−6557. https://doi.org/10.1109/CVPR46437.2021.00648
23. K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, *IEEE Signal Process Lett.*, **23** (2016), 1499−1503. https://doi.org/10.1109/LSP.2016.2603342
24. Y. Wang, J. M. Solomon, Prnet: Self-supervised learning for partial-to-partial registration, in *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, **32** (2019), 8812−8824. Available from: https://proceedings.neurips.cc/paper/2019/file/ebad33b3c9fa1d10327bb55f9e79e2f3-Paper.pdf.
25. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, preprint, arXiv:1409.1556.
26. D. P. Kingma, P. Dhariwal, Glow: Generative flow with invertible 1x1 convolutions, in *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, **31** (2018), 10236−10245. Available from: https://proceedings.neurips.cc/paper/2018/file/d139db6a236200b21cc7f752979132d0-Paper.pdf.
27. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 770−778. https://doi.org/10.1109/CVPR.2016.90
28. C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, Bisenet: Bilateral segmentation network for real-time semantic segmentation, in *Computer Vision – ECCV 2018*, **11217** (2018), 334−349. https://doi.org/10.1007/978-3-030-01261-8_20
29. T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019), 4401−4410. https://doi.org/10.1109/CVPR.2019.00453
30. X. Huang, S. Belongie, Arbitrary style transfer in real-time with adaptive instance normalization, in *2017 IEEE International Conference on Computer Vision (ICCV)*, (2017), 1501−1510. https://doi.org/10.1109/ICCV.2017.167
31. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial nets, in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, **2** (2014), 2672–2680. https://dl.acm.org/doi/10.5555/2969033.2969125

32. J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in *Computer Vision – ECCV 2016*, **9906** (2016), 694−711. https://doi.org/10.1007/978-3-319-46475-6_43
33. D. P. Kingma, J. Ba, Adam: a method for stochastic optimization, preprint, arXiv:1412.6980.
34. Y. Lyu, J. Dong, B. Peng, W. Wang, T. Tan, SOGAN: 3D-aware shadow and occlusion robust GAN for makeup transfer, in *Proceedings of the 29th ACM International Conference on Multimedia*, (2021), 3601−3609. https://doi.org/10.1145/3474085.3475531
35. J. Liao, Y. Yao, L. Yuan, G. Hua, S. B. Kang, Visual attribute transfer through deep image analogy, preprint, arXiv:1705.01088.
36. Q. Gu, G. Wang, M. T. Chiu, Y. W. Tai, C. K. Tang, Ladn: Local adversarial disentangling network for facial makeup and de-makeup, in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, (2019), 10481−10490. https://doi.org/10.1109/ICCV.2019.01058
37. B. Yan, Q. Lin, W. Tan, S. Zhou, Assessing eye aesthetics for automatic multi-reference eye in-painting, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), 13509−13517. https://doi.org/10.1109/CVPR42600.2020.01352

AIMS Press