International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 790-802 www.ijaem.net ISSN: 2395-5252

HAMM: Makeup Migration network with High-frequency Information Retained

Wenjie Fei

[1,2]*Senior engineer, School of media arts and communication, Nanjing University of the arts, Jiangsu province, China.*

Corresponding Author: Wenjie Fei

Submitted: 01-02-2022 Revised: 11-02-2022 Accepted: 14-02-2022

ABSTRACT: The extraction of makeup from different faces and the realization of makeup migration have made great progress in computer vision. However, the existing methods have some limitations: (1) The effect of face makeup migration under different postures is poor; (2) For complex situations such as high-light blush, it cannot be taken into account; (3) It is impossible to realize the migration of complex pattern makeup in the face. In order to solve the problem of poor migration between different facial gestures and expressions, this paper designs a High frequency-assistance Attentive Makeup Morphing module, HAMM, which can better generate more realistic face images by constructing a mapping matrix with attention mechanism for source images and target images. At the same time, in order to cope with the loss of high-frequency information such as high-light blush, High Frequency information Guide structure module (HFG) was introduced into HAMM. HFG can be used to enhance the relative high frequency characteristics in the reference image, so as to alleviate the influence of the change of pixel value in the image on the cosmetic error migration. Then, another Pattern Segment Module branch is introduced to realize the migration of complex patterns. PSM can separate the mask of complex patterns from the reference face image and transfer it to the target face. Experimental results show that our model has achieved advanced results.

KEYWORDS: Makeup migration, Pattern migration, Attention mapping matrix, High frequency information guide structure

I. INTRODUCTION

Human beings are congenitally fascinated by beautiful things. Since the establishment of human society, people have been keen to use various cosmetic tools to render and modify their faces: from powder base, eye makeup, eyebrows, blush, lip makeup, etc. However, limited by the wide variety of cosmetics, different brands, colors, and different ways of use, professional stylists are often required to guide them to select the makeup suitable for their own facial features, which will cost unnecessary time and economic costs. Virtual make-up technology and make-up apps are popular with young people at present, but virtual make-up technology is often limited to face style or filter-like changes, and make-up apps are only guided to realize the make-up recommendation mode. Both cannot migrate the make-up on other people's faces to their own faces to realize pixel-level make-up operation, and users cannot know the effect of the same make-up on their own faces.

In previous studies, Tong [1] proposed a facial makeup migration method similar to physical makeup. The core of this method is to divide the source image and reference image into three layers: facial structure layer, skin detail layer and color layer, and then migrate the makeup information of the reference image to the source image through each layer. The migration process of this method is complex, the processing speed is slow and the time is long. Li [2] decomposes the image into several intrinsic layers, according to the reflection model based on physics, simulates the makeup by manipulating the layer, and finally realizes the facial makeup. This method directly operates the layer according to cosmetic attributes, and does not require face alignment before and after sample makeup in the dataset. However, since this method largely depends on the accuracy of intrinsic layer decomposition, the decomposition error will reduce the quality of migration results. In addition, its processing speed needs to be improved. The traditional makeup migration technology failed to achieve the desired migration effect in performance and quality.

With the development of artificial neural network technology, artificial neural network has

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 790-802 www.ijaem.net ISSN: 2395-5252

realized pixel-level segmentation and classification of images. In face recognition technology, DeepFace[3] improves the face alignment method by selecting the eye center, nose point and three mouth points as the reference points. Jie[4] proposed that the HFE module realized the mining of the high-frequency edge features of the image by the deconvolution method of first down-sampling and then up-sampling. R Ranjan [5] proposed a multi-task learning frame work that can simultaneously detect face global and local key points such as posture angle and smile expression to further improve the analysis ability of local facial features. Toward Open [6] collects the identity information of the source image (key information such as eyes, nose and mouth) and the posture information of the target image through GAN network to generate a new face through confrontational training. Full face recognition technology research provides the possibility for the migration of makeup at the pixel level.

Although the makeup migration method based on convolutional neural network can achieve a certain degree of makeup migration effect, it is still divided from the source image to a specific makeup and then migrated to the target image by layer channel fusion one by one, which is not divorced from the scope of image segmentation. Such makeup migration often has realistic fragmentation and the overall migration effect is not natural enough. Inspired by the GAN network in the generation of fused human faces, the generation of confrontation network can train true and false samples through confrontation to achieve visual reality and face images that do not exist in reality. At present, the face makeup migration method based on generative confrontation network has become a research hotspot in the field of face makeup migration. BeautyGAN [7] is the first method to use GAN for makeup transfer. From the perspective of human visual perception, compared with traditional methods, the transfer effect has been significantly improved. However, BeautyGAN can only deal with face images for simple style makeup transfer. PSGAN [8] can deal with makeup transfer of different postures and expressions. CPM [9] can realize pattern transfer and makeup transfer. However, these methods have some limitations (1) For complex situations such as pattern blush, it cannot be taken into account (2) The robustness of the transfer is greatly reduced when there is a light and dark difference in the face. To solve these problems we are going to start from two aspects.

In view of the situation of complex makeup, we believe that some related work [PSGAN, BeautyGAN, GLOW] is often difficult to deal with the makeup that occupies most of the face such as pattern and blush. Only CPM can complete both light makeup migration and pattern patch tasks. However, due to the limitation of mapping UV space, CPM deals with the migration of face makeup. For different postures and even different face shapes, it is impossible to migrate naturally. We try to learn from the dual-branch idea of CPM, and one branch completes the migration of face makeup with different postures. Another branch completes the pattern segmentation and patching. Experiments show that our model is natural and regular for the migration of complex makeup when the model converges well.

Aiming at the problem of poor face brightness and darkness, the traditional generative adversarial network is easy to identify the dark tone as the gray powder of the makeup. The reason is that the extraction of makeup elements is from the perspective of the whole picture, which is prone to a wide range of makeup similar migration. We believe that as long as improving the positioning accuracy of the makeup area can alleviate this problem, because the change of RGB value in the light and dark area is not more intense than the change of face contour. Since the regions with large changes in image re-frequency domain decomposition are slightly high-frequency regions, we try to use the high-frequency semantic guidance module to guide the migration of the makeup part of the model.

Our contributions in this article are summarized below:

• We developed a new GAN method for makeup migration, which can realize pose/expression, robust and accurate makeup migration according to the high-frequency information of the image.

• We propose an optimized generator bottleneck module that takes more details into account when extracting original and reference images.

• We use the semantic guidance module of high frequency information in the cosmetic matrix extracted from the reference image to reduce the influence of light and darkness on the error migration of cosmetic appearance.

II. RELATED WORK

Makeup migration based on traditional methods. Traditional methods [1,10] focus on image preprocessing techniques, such as coordinate pixel extraction and adjustment [11] or reflection operation [2] , but the use of traditional image magnification, an automatic framework for example-based virtual makeup. e processing techniques for data set integrity [12], the accuracy of layer operation have higher requirements[2]. Supervisory learning method based on convolutional

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 790-802 www.ijaem.net ISSN: 2395-5252

neural network has achieved remarkable results in the field of computer vision. Liu [13] proposed a deep local makeup migration network, and selected the most similar images to the current plain face from the makeup face database, and used the full convolution image segmentation network for face segmentation and extraction of facial features. Finally, the makeup migration of powder (corresponding to face), lip color (corresponding to both lips) and eye shadow (corresponding to both eyes) is completed. Although this method can control the makeup concentration, the overall effect is not natural enough. Wang [14] proposed a new method based on convolutional neural network. Firstly, the feature information of source image and reference image was located and extracted, and the automatic migration of makeup was realized through the makeup migration network and loss function. Huang [15] proposed a multi-path regional fast makeup migration network model. End-to-end face calibration was completed by face key point detection. The loss function of path difference was used to optimize the network according to the makeup characteristics of different facial regions. Finally, the migration results were generated by Poisson fusion and multi-path output. However, all kinds of deep learning frameworks based on convolutional neural network were concentrated in the field of image segmentation and target detection. In the application of makeup migration, the local features of segmented face were used.

Makeup transfer based on GANs. Due to the excellent performance of GANs in the field of unsupervised learning and the support of hardware computing power, the main research of makeup transfer task is focused on the model research based on GANs. The generative adversarial network completes the learning through the game between two neural networks. In the field of image transfer, the generative adversarial network continuously learns the feature distribution of the target image by identifying the true and false samples, so that the source image is generated into an image close to the distribution of the target image, which is more realistic and similar visually. CycleGAN-based[16] model is introduced to transform face-to-face makeup style in an unsupervised way, but CycleGAN can only realize the global domain style migration of images, ignoring the migration effect of local facial features. In order to obtain more realistic results, BeautyGAN [7] proposed a dual-input / dual-output generation confrontation network framework, which combines global domain-level loss and local instance-level loss in the same network, and finally realizes the migration of makeup, especially the makeup of key parts such as eyebrows, eyes and lips. But BeautyGAN learns the facial makeup style as well as the spatial distribution of a specific face, which leads BeatyGAN to apply only to face makeup migration at the same angle without landing value. Beauty GLOW[17] proposed using GLOW framework to decompose makeup and non-makeup components in potential space. LADN [18] combines multiple and overlapping local discriminators for extreme cosmetic transfers, but performs unsatisfactory in fineness. Recently, PRNET [19] proposed a 3D face position mapping UV module. The UV module separated the key information of face position and mapped it to a position information space. The pixel migration of point-to-point was completed in the position information space. Finally, the makeup migration of the face was realized by returning the reflection to the visual layer. The UV module effectively solved the problems of face position and dynamic expression, but it often had the problem of losing edge information. PSGAN [8] uses the makeup extraction network to decompose the makeup of the reference image into two space-aware makeup matrices. Then, note that the makeup deformation module is introduced to specify how the pixel makeup in the source image is deformed from the reference image, which effectively alleviates the problem of edge information loss.

III. FRAMEWORK

On the basis of previous great research, we standardize the makeup migration as an unsupervised learning task of image domain migration, as shown in 3.1. Then we will describe our overall network framework, and introduce our makeup extraction network and makeup deformation module in 3.2 and 3.3, respectively. Finally, in Section 3.4, we will introduce several Losses we use.

3.1. Task Definition

Let X and Y be called the source image domain and the reference image domain. We use the $\{x^n\}_{n=1,\dots N}, x^n \in X$ 和 $\{y^m\}_{m=1,\dots M}, y^m \in Y$ to identify the examples of the two image domains. We do not need paired data sets, and the source image and the reference image can have different identities. We assume that x is sampled from the source image domain by probability ρ_x, and y is sampled from the reference image domain by probability ρ_Y The task of our model is to learn a transfer function G to map x to the $\tilde{x}$ of the reference domain, where $\tilde{x}$ is the makeup style of the reference domain image.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 790-802 www.ijaem.net ISSN: 2395-5252

3.2. Frame

The overall structure of our model in Figure 1 is designed as a double branch structure to handle two tasks, makeup color migration and pattern detail filling. For the color transfer of makeup, our model takes the source image x and the reference image y as input, and the makeup extraction network (MDnet) extracts the simple makeup style of y into γ and β two makeup matrices. Due to the great difference between the source image and the reference image in the head posture and expression, the extracted two matrices cannot be directly applied to the source image$x$. We propose a high-frequency information makeup style deformation module HAMM, which changes these two matrixes into two new matrixes γ' and β'. Considering that the makeup information is high-frequency information in the frequency domain of the image, we use high-frequency information strengthening to make the makeup migration more adaptive and more natural. Then the adaptive makeup matrix γ' and β' are applied to the input of the decoding part, and the makeup transmission is realized by the multiplication and addition of the elements. The output image obtained by the Manet decoding architecture is used to realize the simple makeup color migration. For pattern details filling, we use multi-scale pooling and high-frequency information to guide the segmentation of the detail pattern $L_{ref}^{pattern}$ of the reference image. Under the guidance of adaptive low-dimensional convolution, the output of the decoding module is directly superimposed by channels, so as to achieve controllable and robust makeup migration.

Table 1. Generating Network Encoder Structure

Part	InputShape→OutputShape	Layer reset params
Down-sampling	$(h,w,3+n_c)\rightarrow(h,w,64)$	CONV-(N64,K7*7,S1,P3),IN,ReLU
	$(h,w,64)\rightarrow(\frac{h}{2},\frac{w}{2},128)$	CONV-(N128,K4*4,S2,P1),IN,ReLU
	$(\frac{h}{2},\frac{w}{2},128)\rightarrow(\frac{h}{4},\frac{w}{4},256)$	CONV-(N256,K4*4,S2,P1),IN,ReLU
Bottleneck	$(\frac{h}{4},\frac{w}{4},256)\rightarrow(\frac{h}{4},\frac{w}{4},256)$	Residual Block: CONV-(N256,K3*3,S1,P1),IN,ReLU
	$(\frac{h}{4},\frac{w}{4},256)\rightarrow(\frac{h}{4},\frac{w}{4},256)$	Residual Block: CONV-(N256,K3*3,S1,P1),IN,ReLU
	$(\frac{h}{4},\frac{w}{4},256)\rightarrow(\frac{h}{4},\frac{w}{4},256)$	dResidual Block: CONV-(N256,K3*3,S1,P1),IN,ReLU
	$(\frac{h}{4},\frac{w}{4},256)\rightarrow(\frac{h}{4},\frac{w}{4},256)$	Residual Block: CONV-(N256,K3*3,S1,P1),IN,ReLU
	$(\frac{h}{4},\frac{w}{4},256)\rightarrow(\frac{h}{4},\frac{w}{4},256)$	Residual Block: CONV-(N256,K3*3,S1,P1),IN,ReLU
	$(\frac{h}{4},\frac{w}{4},256)\rightarrow(\frac{h}{4},\frac{w}{4},256)$	Residual Block: CONV-(N256,K3*3,S1,P1),IN,ReLU
Up-sampling	$(\frac{h}{4},\frac{w}{4},256)\rightarrow(\frac{h}{2},\frac{w}{2},128)$	DECONV-(N128,K4*4,S2,P1),IN,ReLU
	$(\frac{h}{2},\frac{w}{2},128)\rightarrow(h,w,64)$	DECONV-(N64,K4*4,S2,P1),IN,ReLU
	$(h,w,64)\rightarrow(h,w,3)$	DECONV-(N3,K7*7,S1,P3),IN,ReLU

Figure 1. Overall framework, the overall framework of the model is a two-branch model, in which Makeup Distill Net and Makeup Apply Net as part of the generative adversarial network are mainly used to refer to the makeup stripping of the face and the common makeup migration from the source image to the target image, and the Detail Segment Module is used to realize the migration of complex patterns.

Makeup distill Network(MDNet). MDNet adopts the encoder structure in StarGAN, which separates features related to simple makeup such as lip glaze color eye shadow color from facial features. We found that the real-time normalization of bottleneck used in the original encoder structure was normalized in a channel, and the mean value of H*W was calculated for stylized migration; because in image stylization, the result depends mainly on an image instance, the whole batch normalization is not suitable for image stylization, so the HW is normalized. It can accelerate the convergence of the model and maintain independence between each image instance. Generate the network encoder structure as shown in the Table 1.

$$\begin{cases} \hat{F}_y = up(\delta(F_y)) & where F_y \in R^{h/2\times w/2\times c/2}, \hat{F}_y \in R^{h\times w\times c} \\ F_h = F_x - \hat{F}_y & where F_x \in R^{h\times w\times c}, F_h \in R^{h\times w\times c} \\ \hat{F}_x = F_x + \gamma \times F_h & where F_h \in R^{H\times W\times C} \end{cases}$$

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 790-802 www.ijaem.net ISSN: 2395-5252

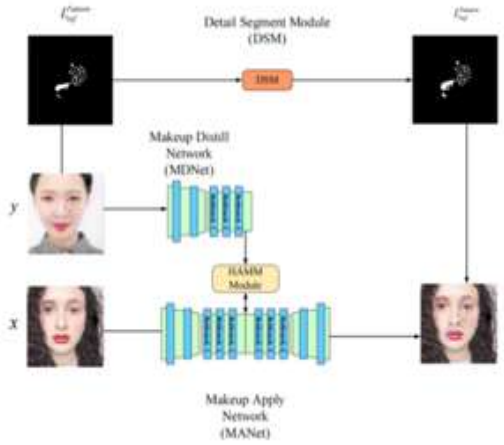

Table 1. High frequency-assistance Attentive makeup morphing module (HAMM) Module

Because the source image and the reference image may present different postures and expressions, the two makeup matrices γ and β extracted by the generator cannot be directly used in the feature map extracted from the source image. We propose that the HAMM module is to realize this function. He calculates a attention matrix $A \in \mathbb{R}^{HW \times HW}$ to represent how the pixel x of a source image is deformed to the reference y domain image. $A_{i,j}$ represents the attention value between the i pixel x_i in the feature graph of x and the j pixel y_j in the feature graph of y. In other words, referring to the relative image position of the makeup part of the image and the relative position of the source image, the attention value between pixel values should be high. In order to describe the relative position mentioned above, we first need to label the facial key points as anchor points. We select 68 facial key points as benchmarks so x_i can be represented as $p_x^i \in \mathbb{R}^{136}$:

$$A_{i,j} = \frac{\exp\left(\left[wv_i, \frac{p_i}{\|p_i\|}\right]^T \left[wv_j, \frac{p_j}{\|p_j\|}\right]\right) \prod(m_x^i = m_y^j)}{\sum_j \exp\left(\left[wv_i, \frac{p_i}{\|p_i\|}\right]^T \left[wv_j, \frac{p_j}{\|p_j\|}\right]\right) \prod(m_x^i = m_y^j)} \quad (2)$$

$$p_x^i = [f(x_i) - f(l_1), f(x_i) - f(l_2), \ldots f(x_i) - f(l_{68}), g(x_i) - g(l_1), g(x_i) - g(l_2), \ldots g(x_i) - g(l_{68})] \quad (1)$$

Where $f(\cdot)$ and $g(\cdot)$ represent the coordinates in the x domain and in the y domain, respectively, and l_i represents the ith coordinate of the key point of face detection. In order to deal with the face with different sizes in the image, we divide the coefficient by its norm when calculating the attention matrix, as shown in Formula 2. In addition, considering the different semantics of position similarity and the visual similarity between pixels, we multiply the specific computational features by a learning weight.

In order to alleviate the influence of the change of light and dark tone on the error transfer of makeup, we incorporate a high frequency semantic guide structure in this module, as shown in Figure 3.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 790-802 www.ijaem.net ISSN: 2395-5252

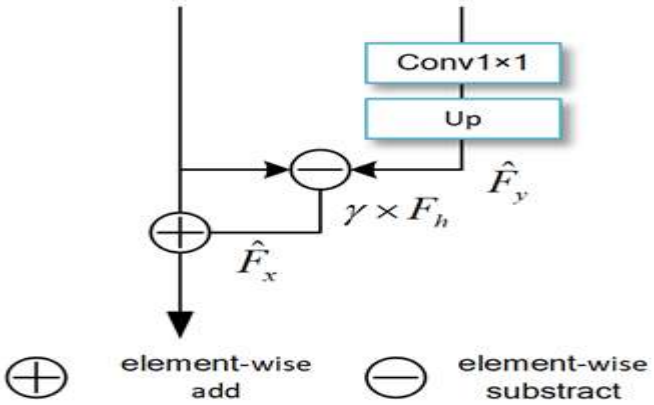

Figure 2. The middle layer image with the same size as the make-up feature map in the reference image domain is obtained by deconvolution of the down-sampled feature map, and then the make-up feature map in the reference image domain is subtracted to obtain the semantic information of relatively high frequency.

As shown in Figure2, its structure is simple and clear. Firstly, its input comes from two parts, one of which is from the makeup feature map in the reference image domain, and the other is from the feature map after down-sampling. These two parts contain different feature information. Under ideal conditions, we consider that the former contains more information than the latter, or the latter is a subset of the former. We amplify the deep feature map by one convolution operation and up-sampling operation to the same size as the shallow feature map, and then subtract the two, so that we can obtain the feature map with relatively high frequency. Although the high-frequency signal is effective, there is also noise. Therefore, we set a learning parameter γ to enable the network to independently adjust the proportion of high-frequency signals. Here, we use an addition operation to overlay our high-frequency feature map onto the original feature map to play a role in high-frequency enhancement. The design idea of this module is to make up for the signal loss caused by down-sampling. The above steps can be expressed as follows:

(3)

Where R represents the real number field, h,w,c represents the height, width and channel number of the feature map, F_x represents the shallow input; F_y represents deep input; δ(·) represents convolution operation for uniform channel number, $up(\cdot)$ represents up sampling operation for uniform feature map size; $\hat{F}_y$ represents the deep input after up-sampling amplification; F_h represents the high-frequency feature map obtained by subtracting F_x and $\hat{F}_y$; γ Represents the weight coefficient; $\hat{F}_x$ represents the final feature map after high frequency enhancement. The enhancement of this module in the frequency domain is equivalent to strengthening the model for the extraction of cosmetic information, making the migration more smooth and natural. In summary, our HAMM module can be expressed as the following structural diagram as follows:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 790-802 www.ijaem.net ISSN: 2395-5252

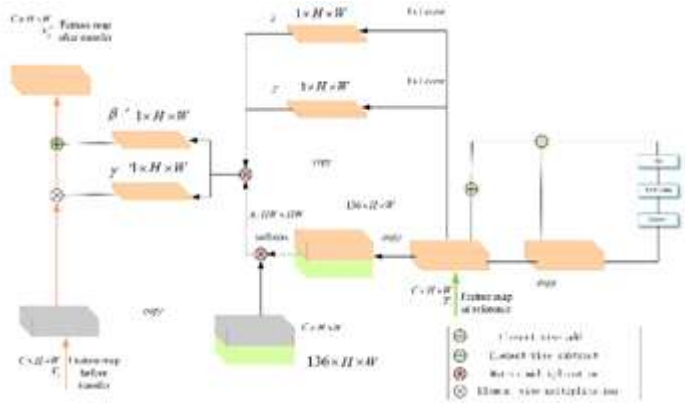

Figure 3. High frequency-assistance Attentive Makeup Morphing module, HAMM.

The above figure shows the HAMM module architecture in our network. The input feature map is the feature map extracted by MDNet. The right side of the figure is the enhancement of our HFE high-frequency information. The enhanced feature map is input into two 1 * 1 convolution modules to extract the makeup matrix, and then the feature map extracted from the source image is processed by softmax operation to output a migrated multi-channel feature map.

Makeup Apply Network (MANet). MANet uses the codec structure mentioned above to align convolution and deconvolution operations. Their structures are very similar but do not share parameters. We also remove the instant normalization to avoid the dressing information extracted from HAMM being normalized by Gaussian distribution to lose information.

Pattern Segment (PSM) Module. Inspired by the dual branch of CPM, we found that there was no migration of large area face painted or pattern in the original PSGAN related work, and we cited a single branch to complete this task as shown in Figure 4.

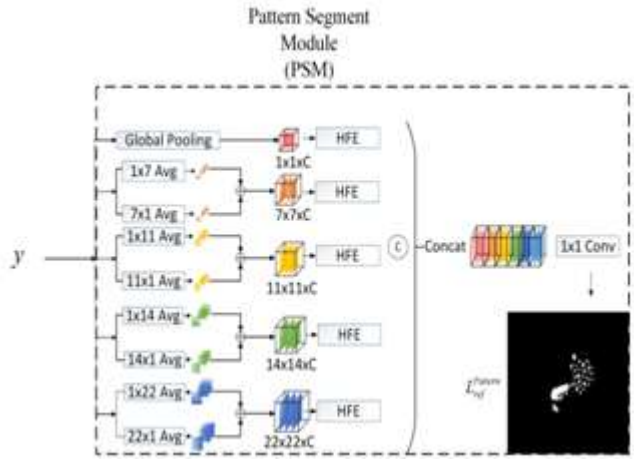

Figure 4. Pattern Segment (PSM) Module.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 790-802 www.ijaem.net ISSN: 2395-5252

We find that when large-scale patterns exist in the face, there will be strong high-frequency information at the edge. We think about whether we can use the high-frequency edge information to guide the pattern migration in the specific implementation. Under the guidance of relevant papers, we find that we can control the proportion of edge information by multi-scale pooling. Under the action of 1X1 convolution, we can separate the pattern mask information. The large-scale face pattern cover migration under this module will not be a problem, and the position information can be adaptively changed under the influence of our neural network parameters. However, the increase in the number of confrontational optimizations caused by the increase in tasks may be different on different machines, so our network is formally formed.

3.3. Objective Function

In this small part, we will introduce the loss function used, and the migration task of makeup migration has a large gap in the demand for the objective loss function used in the neural network.

Adversarial Loss L_{adv} [20]: The generative adversarial network is the backbone of face makeup migration. The purpose of optimizing the adversarial loss function is to improve the ability of the discriminator to identify the shape feature distribution of the key areas of the face (such as eyes and mouth), so that the face image generated by the generator is more realistic in detail. The anti-loss function is calculated as follows:

$$L_D^{adv} = E_{I_{src}}[\lg D(I_{src})] + E_{I_{src},I_{ref}}[\lg(1 - D(G(I_{src}, I_{ref})))]$$
$$L_G^{adv} = E_{I_{src},I_{ref}}[\lg(D(G(I_{src}, I_{ref})))] \quad (4)$$

Where E(∗) represents distribution expectation, G represents generator, D represents discriminator.

Cyclic consistency loss L_{cyc} [21]: Due to the lack of triple data (source image, reference image and migration image), makeup migration is essentially an unsupervised learning problem. In the training process of generating confrontation network, it is difficult to optimize the confrontation target alone. CycleGAN proposes a circular consistency loss function: the target image is regarded as a middleware, and a generator is added after the target image to reconstruct the source image to generate samples, and L_{cyc} is established to ensure that the reconstructed image and the source image maintain consistency in the distribution. CycleGAN believes that imposing constraints on the reconstructed image and the input image can optimize the generator's ability to generate high-quality target images. The definition of circular consistency loss function is as follows:

$$L_{cyc} = E_{I_{src},I_{ref}}\begin{bmatrix} dist(I_{src}^{ref}, I_{src}) \\ +dist(I_{ref}^{rec}, I_{ref}) \end{bmatrix}, \quad (5)$$

Where $(I_{src}^{ref}, I_{src}) = G(G(I_{src}, I_{ref}))$, $dist(\bullet)$ can be L1 norm or L2 norm.

Perception loss L_{per}: In order to ensure the consistency between the face before and after makeup, this paper uses the perception loss function as a branch loss function. The perception loss function can maintain the personal identity information of the source image when transferring the makeup style. The perception loss reduces the difference between the source image and the target image by minimizing the difference between the high-level features extracted by the deep convolution network. The perception loss has been a relatively mature technology in the field of face image. Generally, the pre-trained VGG-16 model [22] on ImageNet can achieve good face fidelity.

The calculation formula of Perception Loss is:

$$L_{per} = E_{I_{src},I_{ref}}[||F_l(I_{src}) - F_l(I_{src}^B)||_2] + E_{I_{src},I_{ref}}\left[||F_l(I_{ref}) - F_l(I_{ref}^A)||_2\right], \quad (6)$$

Where $(I_{src}^B, I_{ref}^A) = G(I_{src}, I_{ref})$, $F_l(x)$ represents the output of layer l of the VGG-16 model.

Makeup loss L_{make} [7]: color capture of key makeup parts such as lips, eyes and faces is the focus of makeup migration. In order to improve the effect of makeup migration at these positions, histogram loss of these

$HM(I_{src}^B \otimes M_{src,i}, I_{ref} \otimes M_{ref,i})$

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 790-802 www.ijaem.net ISSN: 2395-5252

three local colors needs to be established. Histogram matching is obtained in the same facial area of the generated image $I_{src}^{B}$and the reference image I_{ref}, and a remap image is obtained, which constrains the generated image and the reference image to have similar makeup style at the position of M_{item}. M_{item} is a local area obtained by face analysis model, where $item$ represents the set of lips, eyes and faces. The calculation formula of local color histogram loss is:

$$L_{item} = \left\| \begin{matrix} HM(I_{src}^{B} \otimes M_{item}, I_{ref} \otimes M_{item}) \\ -I_{src}^{B} \otimes M_{item} \end{matrix} \right\|_2. \quad (7)$$

Therefore, the total loss of color makeup is calculated as:

$$L_{makeup} = \lambda_1 L_{lips} + \lambda_2 L_{eyes} + \lambda_3 L_{face}. \quad (8)$$

In order to strengthen the learning proportion of lips and eyes, $\lambda_1$is set to 1, λ_2 is set to 1, λ_3 is set to 0.1.

Total loss function: Finally, we confirm the total loss function of the model, which are generator loss and identifier loss:

$$L_D = \lambda_{adv} L_D^{adv}$$
$$L_G = \lambda_{adv} L_G^{adv} + \lambda_{cyc} L_G^{cyc} + \lambda_{per} L_G^{per} + \lambda_{make} L_G^{make}, \quad (9)$$

Whereλ_{adv},λ_{cyc}, λ_{per}, λ_{make} are the hyper parameters of each part loss.

IV. EXPERIMENTS

4.1. Data Collection

In this paper, MT (Makeup Transfer) data set [7] is used as the benchmark model data set. MT is the most widely used data set in most face cosmetic transfer research. The dataset includes 3834 female images, including 1115 vegetarian images and 2719 cosmetic images, including Asians, Europeans and Americans. There are differences in posture, expression and background. The vegetarian images are all bare makeup, and makeup images contain many makeup styles, such as smoke makeup, retro makeup, Korean makeup style and Japanese makeup style. In the process of data cleaning, most of the cosmetic images with exaggerated cosmetic style are deleted, and some images with obvious differences in face pose angles compared with most of the face images are deleted to ensure the purity of the data. Finally, 800 cosmetic images and 480 non-cosmetic images are left, and they are cut into images with a resolution of 256 * 256 pixels to make our facial makeup migration dataset.

In addition, in order to achieve the segmentation of patterns and faces, we select a large number of patterns from the Internet (e. g., affixing flowers, flowers, cinnabar ornaments, facial expression patterns, etc.), and manually cover these patterns as separate channel layers on 480 non-makeup images to form a data set of facial pattern makeup. This part of the data set has the ability to separate the pattern mask from the real makeup.

4.2. Experimental Setting and Details

This paper uses Pytorch as the training framework of the model, PSM uses UNET50 [23] as the pre-training model, and MDNet is fully trained through MT data sets. Model training is discrete in the process, and we use data set distribution to train two main parts: the PSM module responsible for facial pattern migration and the MDNet responsible for overall makeup migration. Evaluation Method of User Perception Evaluation and MS-SSIM [24] (Multiscale Structural Similarity) as Evaluation Model Capability.

PSM module of facial pattern migration: this branch uses CPM-Synt-1 data set and uses supervised learning to train. Each training image has a pattern segmentation mask, the size is 256×256. We use the UNet structure of Resnet-18 as the pre-training encoder. Since the original segmentation mask is nonbinary, we use sigmoid as activation function. After 200 stages of training, the model is batched to 4, and the Adam [25] optimizer is used. The learning rate is set to 0.0002. In order to solve the problem of local oscillation, the learning rate

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 790-802 www.ijaem.net ISSN: 2395-5252

attenuation with step length of 10 and attenuation rate of 0.9 is set.

MDNet of overall makeup migration: this branch uses MT data set and uses unsupervised learning to train. In order to complete unsupervised learning, in each iteration, a makeup image and a non-makeup image are randomly selected to form a pair of exchange images. Similarly, we cut the image of MT dataset and calibrate it to 256 × 256 standard format, and then generate an image instance by StarGAN[26] encoder. In addition, the high-frequency enhancement of the image instance through the HAMM module is designed to enhance the learning of high-frequency information in human face makeup (such as high-frequency features such as lips, blush and eyebrows distinguished from most skin pixels). For the loss function of makeup migration, this paper sets the following hyper parameters to better train the model. The weights of each loss component are as follows: λ_{avd}=1, λ_{cyc}=10, λ_{per}=0.005 and λ_{hist}=1. The weight of histogram matching area is: λ_{face}= 0.1, λ_{eyes} =1, λ_{lips} =1. Training to generate adversarial networks generates a large number of copies that take up GPU resources, setting the batch size to 1. We use the Adam optimizer with learning rate of 0.0002 to train the network until convergence.

Evaluation method

User perception evaluation: User perception evaluation is a subjective evaluation index of style transfer effect. This paper provides a set of (original picture, cosmetic transfer picture, reference picture) generated by five cosmetic transfer models including BeatyGAN, LADN, PSGAN and the proposed model. Ten professionals are randomly selected for subjective scoring. Professionals need to makeup migration after the image quality, realism and pattern migration retention factors as the basis. The makeup transfer effect of these sets is subjectively scored in the range of 0-100.

MS-SSIM: Multiscale structural similarity (MS-SSIM) is a method to calculate the similarity of image feature distribution from multiple scales, and its calculation formula is:

$$\mathrm{MS-SSIM} = [\mathrm{L}_M(\mathrm{X},\mathrm{Y})]^{\alpha \mathrm{M}} \prod_{j=1}^{N} [c_j(X,Y)]^{\beta j} [s_j(X,Y)]^{\gamma j} \quad (10)$$

L(X,Y) is brightness contrast factor, C(X,Y)is contrast factor, S(X,Y) is structure contrast factor, α, β and γ are the weight of each component. When M = 1, it represents the original image; when M = 2, it represents that the original image is reduced by half, and so on. The purpose of MS-SSIM is to measure the similarity between illumination and contrast between the generated image and the original image.

4.3. Qualitative experiments

We compare our model with existing published source works, including DMT [27], BeautyGAN[7], LADN[18] and PSGAN[8] . We only extract some of the work that published the source code without worrying about some of the new published work and the effect of SOTA replication. We present some qualitative results here to test the ability of the model to enhance the effect of high-frequency makeup and realize the migration of facial patterns. Due to equipment reasons, the effects of some models cannot be completely repeated. We try to take the most representative correlation comparison in the same training times and time.

Figure 5. Migration effect of our model in different reference domains.

As shown in the figure, our model can better achieve the task of cosmetic migration. As shown in the figure, our model can well capture the color of the lips, similar to the most advanced BeautyGAN. In addition, due to the high-frequency enhancement scheme based on HAMM, our method can successfully transfer blush. We randomly selected a source image in the MT dataset and performed facial makeup migration. The representative results are shown in the second line of Figure 5. Although not seen during training, our network can capture its patterns well and transmit them to the output.

Figure 6. The comparison between our model and other models in the migration effect of makeup.

Next we will compare our model with the one mentioned above. As shown in the Figure 6, our model can well complete the migration of bottom makeup and pattern. When the migration of makeup

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 790-802 www.ijaem.net ISSN: 2395-5252

is complex, the robustness of DMT model which uses decoupling representation to solve the migration of makeup is undoubtedly revealed. At the same time, BeautyGAN has a good effect on color transfer such as lip makeup but can't complete local interpolation makeup transfer at the same time and he doesn't fully take into account the makeup color of the reference image. LADN uses multiple overlapping local adversarial discriminators to achieve local detail migration between facial images. But the face fit is not good. PSGAN cannot achieve local interpolation makeup migration, which is also its limitation. Our model can be more natural and robust migration makeup

Dataset	DMT	BeatyGAN	LADN	PSGAN	OurModel
MT	79.45	85.664	57.257	83.445	**89.246**

Table 2. User perception evaluation

Dataset	DMT	BeatyGAN	LADN	PSGAN	OurModel
MT	0.795	0.760	0.472	0.753	**0.840**

Table 3. Multiscale structural similarity, MS-SSIM

The comparison table of user perception evaluation shows the qualitative results of MT data set. Each element represents the average interval of user evaluation is 0 – 100. It can be seen from the table that our model has achieved high scores in this link. Similarly, the images generated by our model in the MS-SSIM table also reach the highest similarity.

3.4. Ablation Studies

In order to prove the effectiveness of the two proposed modules, we need to perform ablation experiments on our modules alone.

The first is the high-frequency information guidance module proposed by us. Without the details segmentation module, we test whether our module is more natural and robust to the migration of makeup.

Figure 7. For the ablation experiment of high frequency information guide module, the hyperparam-eter γ represents the different positions of initialization

In specific experiments, we find that when the hyper-parameter γ of the high-frequency information guidance module is given to initialize the parameter, the final convergence effect of different parameters is different. Even if it is a learning parameter but initializes to a small number, it cannot learn a good interval. So we finally try to give it a golden section ratio, so it's magical to have this effect. It can be clearly seen from the figure that when a high-frequency information guidance module is added, he is highly sensitive to the region with strong changes, and can migrate the makeup very naturally.

Table 4. Influence of User Perception Evaluation on User Feedback in Different Hyperparameter Settings and without High Frequency Information Guidance Module

Dataset	None	$\gamma = 0.1$	$\gamma = 0.618$
MT	0.756	0.760	0.840

Table 5. The influence of different super-parameters and no high-frequency modules on the final SSIM calculation values

Dataset	None	$\gamma = 0.1$	$\gamma = 0.618$
MT	81.659	80.740	90.579

As can be seen from the table, when γ = 0.618, the score of user perception evaluation and MS-SSIM are best.

Next, we will do ablation experiments for our model's pattern detail segmentation patch module.

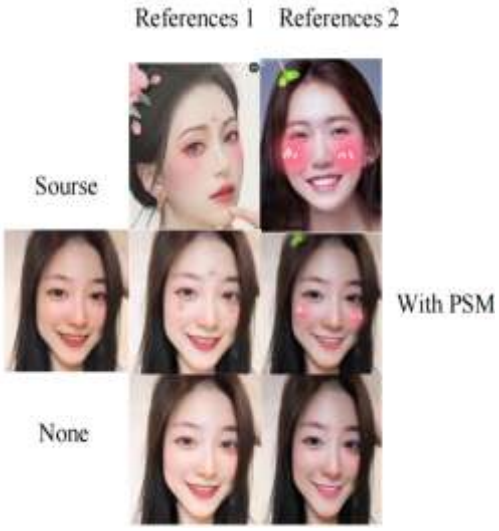

Figure 8. PSM Module Ablation Experiment

As shown above, we find that when the reference image has a large range of patterns, the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 2 Feb 2022, pp: 790-802 www.ijaem.net ISSN: 2395-5252

segmentation of the PSM module and the patch can complete this task well. In the end, we put out more experimental results. Considering the problem of portrait rights, we need to carefully select images, and finally select more reliable images for migration.

Figure 9. Our model migrates to other people's faces where the first three in each row are reference images and the last three are respective effects.

CONCLUSION

In this paper, based on previous studies, we propose high-frequency information guidance module and detail pattern makeup migration module, which can also converge to better results when hardware resources are not very sufficient. In the specific experiment, we found the influence of the hyper-parameter presupposition of high-frequency information on the results and the dependence of the pattern transfer module on the posture. It was found that when the postures of the reference image and the original image were completely mismatched, the pattern patch would have a great mis-transfer situation. Our makeup transfer method was suitable for use under the condition that the postures of the given reference image and the original image did not change in a wide range, but the face was dark and the bottom makeup distribution was extremely uneven. We hope that the follow-up work can solve the problem of pattern error migration.

Conflicts of Interest: The authors declare no conflict of interest.

SOME OF THE ADVANAGES FROM THE ABOVE RESULTS

a) We developed a new GAN method for makeup migration, which can realize pose/expression, robust and accurate makeup migration according to the high-frequency information of the image.

b) We propose an optimized generator bottleneck module that takes more details into account when extracting original and reference images.

c) We use the semantic guidance module of high frequency information in the cosmetic matrix extracted from the reference image to reduce the influence of light and darkness on the error migration of cosmetic appearance.

REFERENCES

[1]. Tong, W.S.; Tang, C.K.; Brown, M.S.; Xu, Y.Q. Example-based cosmetic transfer. 15th Pacific Conference on Computer Graphics and Applications (PG'07). IEEE, 2007, pp. 211–218.

[2]. Li, C.; Zhou, K.; Lin, S. Simulating makeup through physics-based manipulation of intrinsic image layers. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4621–4629.

[3]. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1701–1708.

[4]. Yang, J.; Fan, J.; Wang, Y.; Wang, Y.; Gan, W.; Liu, L.; Wu, W. Hierarchical feature embedding for attribute recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 13055–13064.

[5]. Ranjan, R.; Sankaranarayanan, S.; Castillo, C.D.; Chellappa, R. An all-in-one convolutional neural network for face analysis. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). IEEE, 2017, pp. 17–24.

[6]. Bao, J.; Chen, D.; Wen, F.; Li, H.; Hua, G. Towards open-set identity preserving face synthesis. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6713–6722.

[7]. Li, T.; Qian, R.; Dong, C.; Liu, S.; Yan, Q.; Zhu, W.; Lin, L. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. Proceedings of the 26th ACM international conference on Multimedia, 2018, pp. 645–653.

[8]. Liu, Q.; Zhou, H.; Xu, Q.; Liu, X.; Wang, Y. PSGAN: A generative adversarial network for remote sensing image pan-sharpening. IEEE Transactions on Geoscience and Remote Sensing 2020.

[9]. Nguyen, T.; Tran, A.T.; Hoai, M. Lipstick Ain't Enough: Beyond Color Matching for In-the-Wild Makeup Transfer. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13305–13314.

[10]. Hertzmann, A.; Jacobs, C.E.; Oliver, N.; Curless, B.; Salesin, D.H. Image analogies. Proceedings of the 28th annual conference on

Computer graphics and interactive techniques, 2001, pp. 327–340.
[11]. Xu, L.; Du, Y.; Zhang, Y. An automatic framework for example-based virtual makeup. 2013 IEEE International Conference on Image Processing. IEEE, 2013, pp. 3206–3210.
[12]. Scherbaum, K.; Ritschel, T.; Hullin, M.; Thormählen, T.; Blanz, V.; Seidel, H.P. Computer-suggested facial makeup. Computer Graphics Forum. Wiley Online Library, 2011, Vol. 30, pp. 485–492.
[13]. Liu, S.; Ou, X.; Qian, R.; Wang, W.; Cao, X. Makeup like a superstar: Deep localized makeup transfer network. arXiv preprint arXiv:1604.07102 2016.
[14]. Wang, X.; Wang, K.; Lian, S. A survey on face data augmentation for the training of deep neural networks. Neural computing and applications 2020, pp. 1–29.
[15]. Ma, X.; Zhang, F.; Wei, H.; Xu, L. Deep learning method for makeup style transfer: A survey. Cognitive Robotics 2021, 1, 182–187.
[16]. Chang, H.; Lu, J.; Yu, F.; Finkelstein, A. Pairedcyclegan: Asymmetric style transfer for applying and removing makeup. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 40–48.
[17]. Chen, H.J.; Hui, K.M.; Wang, S.Y.; Tsao, L.W.; Shuai, H.H.; Cheng, W.H. Beautyglow: On-demand makeup transfer framework with reversible generative network. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10042–10050.
[18]. Gu, Q.; Wang, G.; Chiu, M.T.; Tai, Y.W.; Tang, C.K. Ladn: Local adversarial disentangling network for facial makeup and de-makeup. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 10481–10490.
[19]. Feng, Y.; Wu, F.; Shao, X.; Wang, Y.; Zhou, X. Joint 3d face reconstruction and dense alignment with position map regression network. Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 534–551.
[20]. Liu, M.Y.; Tuzel, O. Coupled generative adversarial networks. Advances in neural information processing systems 2016, 29, 469–477.
[21]. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE international conference on computer vision, 2017, pp. 2223–2232.
[22]. Liu, Z.; Wu, J.; Fu, L.; Majeed, Y.; Feng, Y.; Li, R.; Cui, Y. Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion. IEEE Access 2019, 8, 2327–2336.
[23]. Li, X.; Chen, H.; Qi, X.; Dou, Q.; Fu, C.W.; Heng, P.A. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE transactions on medical imaging 2018, 37, 2663–2674.
[24]. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 2004, 13, 600–612.
[25]. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 2014.
[26]. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8789–8797.
[27]. Zhang, H.; Chen, W.; He, H.; Jin, Y. Disentangled makeup transfer with generative adversarial network. arXiv preprint arXiv:1907.01144 2019.