

# Makeup transfer: A review

Feng He<sup>1,2,3,4</sup>  | Kai Bai<sup>1,2,3</sup> | Yixin Zong<sup>4</sup> | Yuan Zhou<sup>5</sup> | Yimai Jing<sup>6</sup> |  
Guoqiang Wu<sup>7,8</sup> | Chen Wang<sup>9</sup>

<sup>1</sup>School of Computer Science, Yangtze University, Jingzhou, China

<sup>2</sup>Xi'an Key Laboratory of Tight Oil (Shale Oil) Development (Xi'an Shiyou University), Xi'an, China

<sup>3</sup>Cooperative Innovation Center of Unconventional Oil and Gas, Yangtze University (Ministry of Education & Hubei Province), Wuhan, China

<sup>4</sup>Bureau of Frontier Sciences and Education Chinese Academy of Science, Beijing, China

<sup>5</sup>School of Information and Mathematics, Yangtze University, Jingzhou, China

<sup>6</sup>School of Geosciences, Yangtze University, Jingzhou, China

<sup>7</sup>China Aerospace Science and Technology Corporation, Intelligent Unmanned System Overall Technology Research and Development Center, Beijing, China

<sup>8</sup>Beijing Institute of Technology School of Mechatronical Engineering, Beijing, China

<sup>9</sup>School of Computer Science and Engineering, Beihang University, Beijing, China

## Correspondence

Guoqiang Wu, China Aerospace Science and Technology Corporation, Intelligent Unmanned System Overall Technology Research and Development Center, No.1, Fengying East Road, Haidian District, Beijing, 100094, China.

Email: [18601239845@126.com](mailto:18601239845@126.com)

## Funding information

the Scientific Research Projects of the Hubei Provincial Department of Education, Grant/Award Number: D20201304; Open Fund of Xi'an Key Laboratory of Tight oil Development, Grant/Award Number: XSTS-202101; the National Natural Science Foundation of China under Grant, Grant/Award Numbers: No. 51974036, No. 62006028; the Open Foundation of Cooperative Innovation Center of Unconventional Oil and Gas, Yangtze University, Grant/Award Number: UOG2022-06

## Abstract

Makeup transfer (MT) aims to transfer the makeup style from a given reference makeup face image to a source image while preserving face identity and background information. In recent years, MT has attracted the attention of many scholars, and it has a wide range of application prospects and research value. Since then, many methods have been proposed to accomplish MT, most of which are based on Generative Adversarial Network methods. A taxonomy of existing algorithms in the field of MT is first proposed. Then, evaluation methods are proposed, existing methods are analysed, and existing datasets are introduced. This paper finally discusses the current problems in the field of MT and the trend of future research.

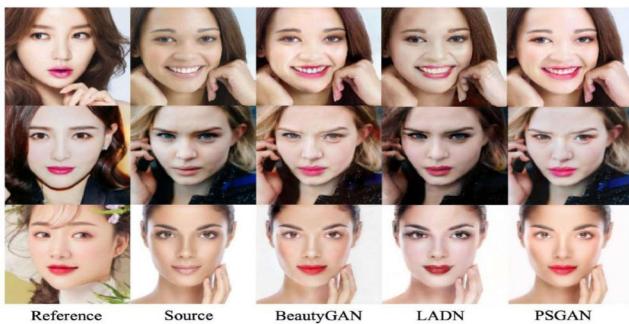
## 1 | INTRODUCTION

Facial makeup transfer (MT) aims to transfer makeup style from a given reference face image to another non-makeup face image while preserving the facial identity. Figure 1 shows the classic MT result. This is an interesting but challenging task. There is a need to pay attention to the process of MT: lipstick, eye shadow, foundation, blush etc. In addition, due to limited resources, it is impossible to really try all the makeup style in a short period of time. Therefore, MT has received widespread attention and has become a recent research hotspot.

In recent years, deep learning has achieved good results in many fields. The proposed Tensorflow deep learning computing framework has further advanced the field of MT [1]. For example, pedestrian re-identification [2–7], face attribute recognition [8–11], mechanical device diagnosis [12], hyperspectral image classification [13–17], style transfer [18–23], smart grid [24, 25] etc. In addition, deep learning has also achieved good results in the field of MT, and currently, MT methods based on generative adversarial networks have become the mainstream methods in the field of makeup change. Style transfer is somewhat similar to the MT method, where the main concern of MT is the retention of identity

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *IET Computer Vision* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.



**FIGURE 1** The makeup transfer (MT) aims to transfer the makeup from a reference image to a source image. This result shows the effect of three classic MT methods

information and the overall effect after makeup application, while the style transfer method is more concerned with synthetic textures [26–29]. There are also methods for style transfer that generate non-realistic synthetic artwork [30–32]. To generate images with realism, some methods with structure preservation and style capture are introduced [33, 34]. InforGAN [35] aims to learn interpretability in the potential space based on information regularisation. Some scholars [36–39] have tried to combine makeup transfers with recommender systems to meet users' choices of makeup. To increase the realism of the generated images, some scholars [40–43] achieve it by rendering locally.

Many exploratory studies [44–47] have been done on MT in the beginning. Ojima et al. [48] use optical properties to simulate the application of cosmetics to the human face, which makes the skin appear smoother and more even. Leyvand et al. use [49] post-modelling and fine-tuning of faces to improve the attractiveness score of faces while guaranteeing strong, unmistakable similarity to the original image. Guo et al. [50] proposed a face image-based makeup method, which is similar to the physical makeup method. First, the face image is divided into facial structure layer, skin detail layer, and colour layer. Although this method is effective, the process is cumbersome and the processing time is long. Scherbaum et al. [51] proposed an MT method based on 3D synthesis and recommended a computer-suggested makeup. Li et al. [52] proposed a new MT method by dividing the face image into reflective layers. After the reflective layer is obtained, it can be calculated through a physics-based reflective model, which can be effectively converted to the equivalent of a cosmetic product. This method reduces a lot of computation compared to the previous rendering, but the effect is limited.

With the advent of deep learning, some scholars [53–55] have used Convolutional Neural Networks (CNN) to accomplish MT task. Liu et. al [56] propose a novel Deep Localised Makeup Transfer Network to intelligently recommend the most suitable makeup for a female and synthesise the make up on her face. This method is an intensity-controllable MT method. First of all, it is necessary to recommend makeup, which is accomplished by calculating the Euclidean distance between the depth feature of the non-makeup picture and the

makeup picture. Then, use Fully Convolution Network [57] for face parsing. Finally, the eye shadow, lips and face are transformed into makeup. Since the MT is carried out locally and lacks the balance of the overall makeup, the final transfer effect is limited.

The traditional MT methods and the MT methods [58–60] based on CNN [61] only consider the effect of the local and lack the consideration of the overall makeup. Although the final effect of MT can be achieved, the overall effect is relatively limited and looks unnatural. With the development of deep learning, more and more methods have been proposed to solve various problems. After the generative adversarial network (GAN) [62] was proposed, it has been widely used in the field of computer vision because it can not only avoid more complex calculations but also generate higher-quality images. At present, in the field of MT, the method based on GAN has become the mainstream method to solve the problem of MT [63].

The organisation of this paper is as follows. First, in the second section, we classify the GAN-based MT methods according to the principles of the methods. Then, in the third part, we introduce the commonly used loss function, which plays a crucial role in the effect of the final MT. In the fourth part, we introduce the commonly used datasets. In the fifth part, we introduced the commonly used evaluation indicators, the mainstream evaluation indicators, mainly divided into two categories, that is, qualitative evaluation indicators and quantitative evaluation indicators. In the sixth part, we point out the possible future directions of development in the field of MT. In the seventh part, we summarise the paper and classic MT methods.

In this paper, we aim to provide an overview of current advances in makeup transfer. Our contributions are threefold. First, we investigate, classify and summarise recent advances in the field of MT. Second, we summarise and evaluate existing methods. Third, we summarise current challenges in this field and propose directions on how to deal with them in future work.

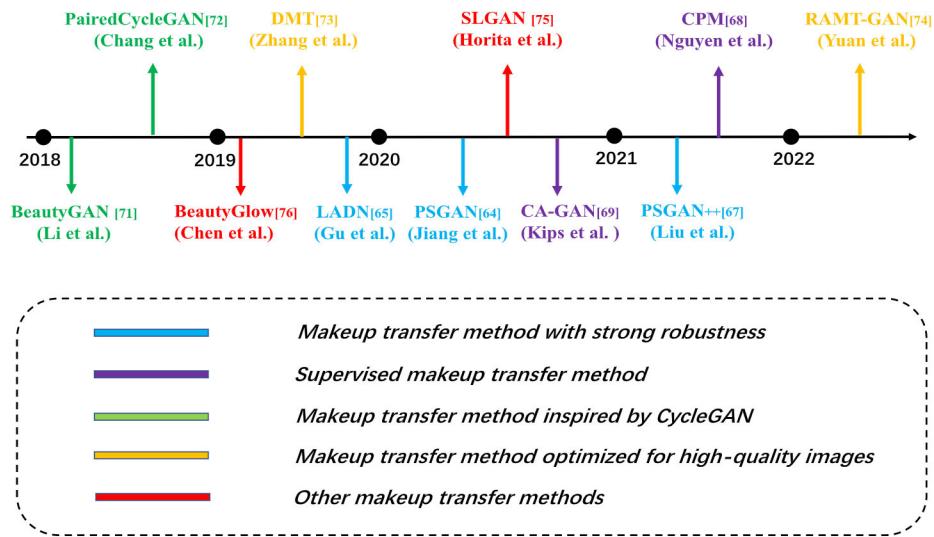
## 2 | MAKEUP TRANSFER METHOD BASED ON GENERATIVE ADVERSARIAL NETWORK

Currently, GAN-based MT methods have become the main methods in the field of MT by their ability to output high-quality images and maintain the original identity information. In this paper, these methods are divided into five categories according to the different methods. The first category is MT methods with strong robustness, which are optimised for large expressions and poses and can still achieve great results in the case of poor expressions and head poses. The second category is supervised MT methods, which are not popular at present, because highly aligned image datasets before and after makeup are hard to collect. The third category is CycleGAN-inspired MT methods, many of which are inspired by CycleGAN to do the work. These methods

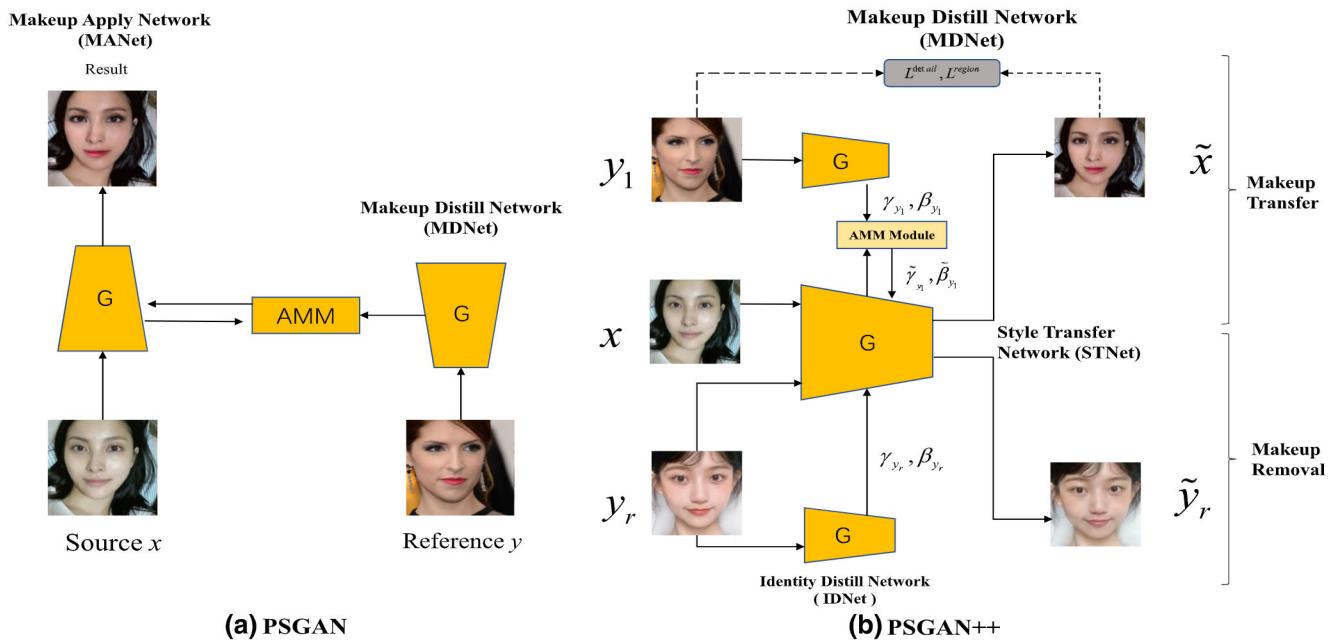
are unsupervised learning methods without a before and after makeup dataset, introducing two functions at the same time, one for makeup application and one for de-makeup. The fourth category is the MT methods optimised for high-quality images, which now have a better performance in terms of image quality compared to the other methods. The fifth category is other MT methods. As shown in Figure 2, we have summarised the important models in recent years in chronological order.

## 2.1 | Makeup transfer method with strong robustness

An MT method with strong robustness means that such methods are optimised for large expressions and postures; even in the case of large differences in facial expressions and poses, good results can be achieved. Representative works are LADN, PSGAN, PSGAN++ etc. Figure 3 shows the network framework diagram of PSGAN and PSGAN++.



**FIGURE 2** Timeline of representative work in the field of makeup transfer (MT). In the past 10 years, the research about makeup transfer has developed rapidly. Some methods are very instructive. For example, BeautyGAN applies pixel-level histogram losses on local regions. Moreover, it introduces cycle consistency and perceptual loss to preserve identify information and generate high-quality images



**FIGURE 3** Framework diagram of the makeup transfer (MT) method with strong robustness. PSGAN and PSGAN++ all belong to the MT method with robustness. PSGAN is the first method to achieve shade-controllable, partial and pose/expression robust in MT field. PSGAN++ aims to address the issue that most make up methods cannot deal with the face images with big pose and expression difference

### 2.1.1 | PSGAN

Large pose and expression MT has been a challenging problem until now. To address this problem, the authors of Ref. [64] propose a pose and expression robust spatially aware GAN (PSGAN). First, it separates the makeup of the reference image into two spatially aware makeup matrices using a compensation network. Then, the Attentive Makeup Morphing (AMM) module is introduced to help the pixels of the source image to be composed. Although, PSGAN produces partial colour difference when performing MT. PSGAN is the first MT model that can simultaneously achieve local, shade-controllable, and pose/expression robustness. PSGAN mainly uses adversarial loss, cyclic consistency loss, perceptual loss, and makeup loss to generate the image for constraint. These four losses are introduced in the third part of this paper. The authors introduce a new makeup transform dataset, Makeup-Wild, to better evaluate the model.

### 2.1.2 | LADN

The authors of Ref. [65] propose a local adversarial disentangling network for makeup application and removal. The authors merge local adversarial discriminators into the domain translation network for makeup application and disentanglement. Also, to ensure cross-cycle consistency, the potential makeup vectors need to be separated from the individual face images. Then, by increasing the number and overlapping local discriminators, a complex makeup style with high-frequency details can be achieved while preserving the original facial and identity information. LADN introduces an asymmetric loss. In Section 3, the asymmetric loss is introduced. By introducing this loss, the model can seamlessly transfer and remove dramatic makeup styles. LADN is an intensity-controlled MT method. Most previous methods only consider simple makeup styles. LADN [65] is the first MT method that considers face painting, blush, jewellery etc. However, the results are not particularly good. The authors also disclose a dataset (LADN Makeup).

### 2.1.3 | SOGAN

When there are shadows and occlusions in the face images, most makeup transfer methods cannot achieve great results. The authors of Ref. [66] proposed a new MT method called 3D-Aware Shadow and Occlusion Robust GAN (SOGAN). First, given the source and the reference faces, the 3D face model is fitted. The fitted model is decomposed into shape and texture, and after that, the texture branch is processed separately. In the texture branch, the authors map the texture into the UV space and apply the makeup by designing a UV tressurer. In addition, a Flip Attention Module is proposed to remove unwanted shadows and occlusions. At the same time, the authors introduce a Makeup Transfer Module to achieve precise makeup application. Finally, the authors illustrate the

superiority of SOGAN compared to other methods through qualitative and quantitative experiments.

### 2.1.4 | PSGAN++

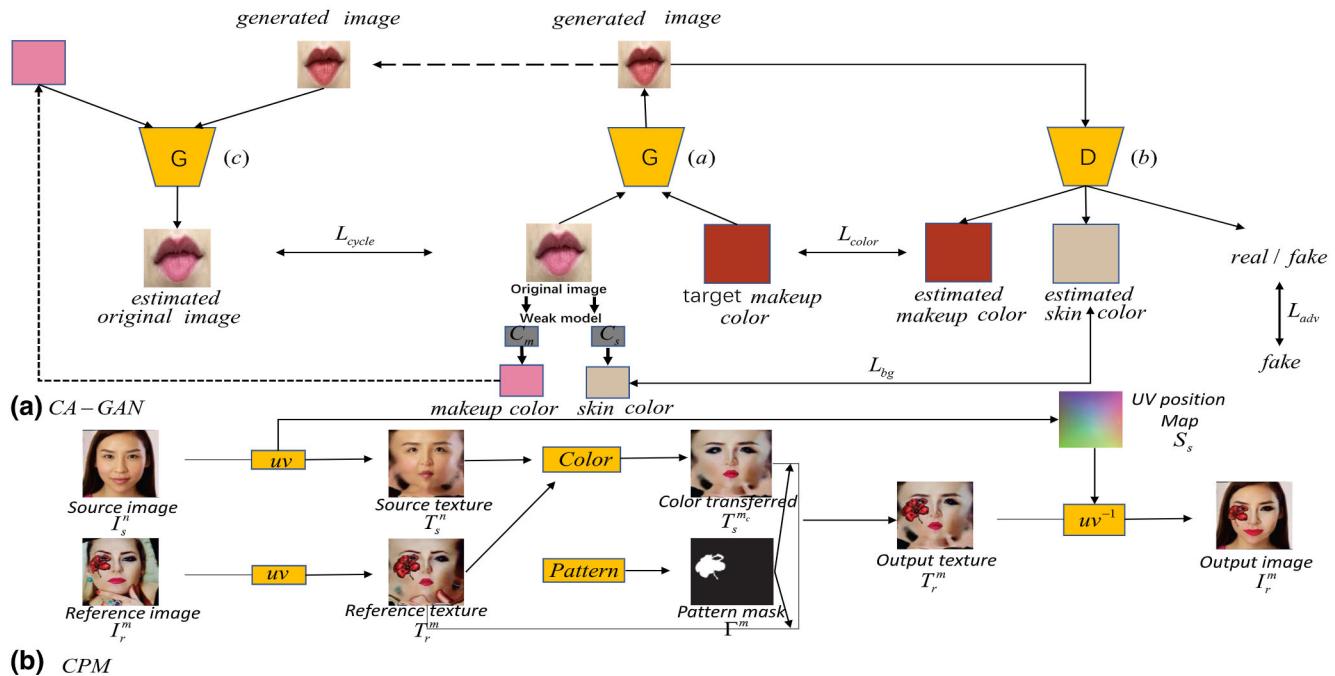
To address the problem that most models have difficulty controlling makeup intensity, the authors of Ref. [67] propose a pose and expression robust spatially aware GAN (abbreviated as PSGAN++). PSGAN++ is capable of performing detail-preserving makeup transfers and effective de-makeup. For makeup application, PSGAN++ uses Makeup Distill Network to propose makeup information, which is embedded in a spatially aware makeup matrix. The authors also devised an AMM module that specifies how the makeup in the source image is morphed from the reference image and a makeup detail loss to supervise the model within the selected makeup. For de-makeup, the authors introduce an Identity Distill Network for embedding the identity information into the identity matrix. Finally, the resulting makeup/identity matrix is sent to a Style Transfer Network for the final makeup application or removal. Also, to evaluate the effectiveness of the model, the authors collected an image dataset MT in the Wild (MT-Wild) and a Makeup Transfer High-Resolution dataset containing various poses and expressions. PSGAN++ is a makeup and de-makeup method that enables pose and expression robustness, localisation, controlled makeup level, and maintenance of details.

## 2.2 | Supervised makeup transfer method

These types of MT methods are often weakly supervised. These methods can eventually achieve good results with human adjustment, but their development is slow due to the lack of paired datasets for MT. The network framework diagram of CPM [68] and CA-GAN [69] is shown in Figure 4.

### 2.2.1 | CPM

The authors of Ref. [68] propose that the MT should include colour transfer and pattern addition. To achieve this, the authors propose a holistic approach that converts colours and patterns from the reference image to the source image. Until then, most approaches have ignored the pattern, focussing only on colour transfer. To fill this gap, a holistic framework for MT is proposed that can handle all makeup transfers. It consists of an improved colour transfer branch and a new pattern transfer branch to learn all makeup attributes, including colour, pattern, texture, and position. To train and evaluate the proposed model, the authors introduce a new MT dataset that contains extreme makeup styles, which are not previously considered. In the colour transfer branch, the network structure is similar to CycleGAN, while the histogram loss proposed by BeautyGAN is used. In the third part of this paper, the histogram loss is introduced. On the colour branch, a



**FIGURE 4** Framework diagram of the supervised approach to makeup transfer (MT). CA-GAN introduces a quantitative analysis of MT performance. CPM designs a novel architecture that combines colour transfer and pattern addition. Moreover, it can work on different makeup degree face images

supervised approach is used to propose makeup patterns. Noticeably, unlike previous methods, both their branches work on warped faces in UV space, thus discarding the discrepancy. The results of the two branches are fused to generate the desired output. It can overcome the limitations of LADN and can handle more complex makeup patterns and have some robustness to head pose.

## 2.2.2 | CA-GAN

To achieve colour-controlled MT, the authors of Ref. [69] propose a GAN for controllable MT. The model is based on a colour regression loss combined with a novel background consistency loss. The model preserves the colour of non-target object attributes. CA-GAN does not require images with colour labels for training, as it can be trained by weak supervision. A new loss function is introduced in order to achieve control over the colour. The authors also propose a new dataset, the social media dataset, with a wider variety of skin tones, facial poses, and makeup colours than the previous dataset, MT. It is worth noting that the authors collected a more controlled dataset called lipstick dataset in order to better evaluate the model.

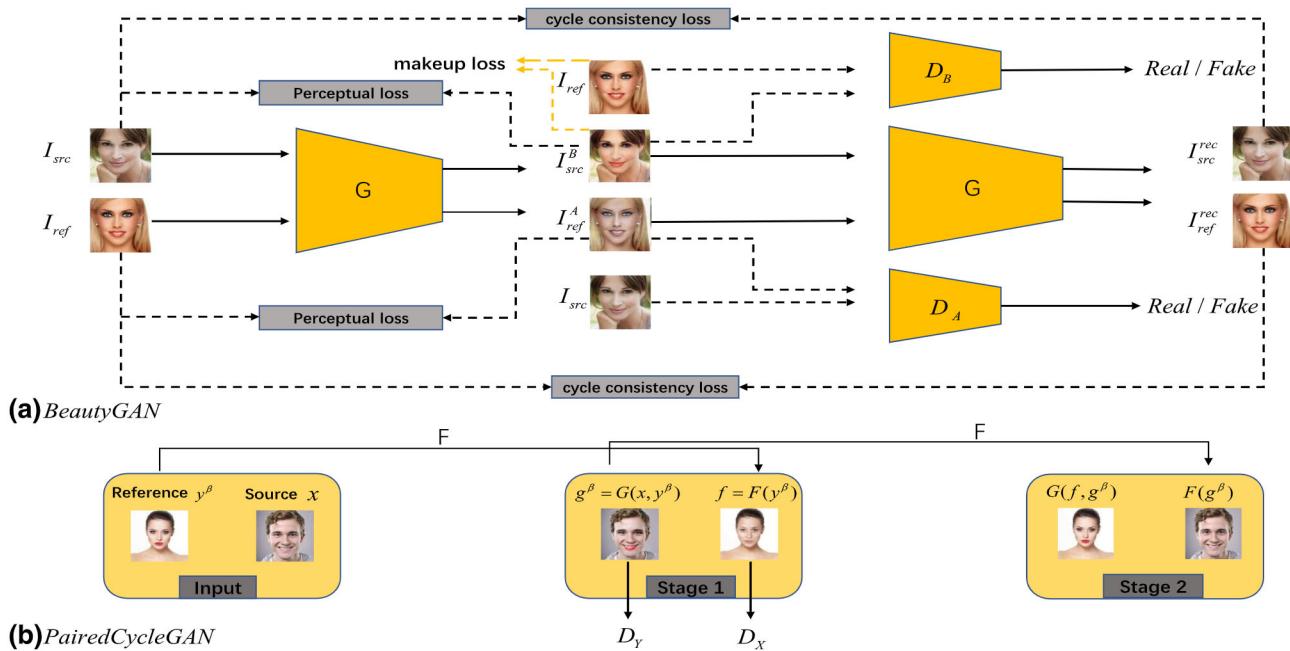
## 2.3 | Makeup transfer method inspired by CycleGAN

In the past few years, CycleGAN [70] has worked well in many tasks. In the field of MT, many methods are implemented

based on CycleGAN's framework in 2D image space. With this type of methods, two functions are usually trained at the same time, one for applying makeup and one for removing it. The network framework diagram of BeautyGAN and Pair-edcycleGAN is shown in Figure 5.

### 2.3.1 | BeautyGAN

Before BeautyGAN was proposed, automatic MT methods could be roughly divided into two categories: the first was traditional image processing methods with image gradient editing and physics-based operations, and the second was deep learning-based methods, and these were usually based on deep neural networks. The previous methods are basically simple combinations of make-up styles, which look unnatural overall and often have more obvious flaws at the combination. To address this problem, the authors of Ref. [72] proposed BeautyGAN. In makeup transfer relative to the previous style transfer, there were two main differences. The first one is a domain style transfer, while MT requires instance-level transfer. Therefore, using CycleGAN cannot achieve good results. The second one is that MT requires a transfer method with higher accuracy. In other words, it is not only necessary to look harmonious on the whole but also to achieve good results in the local area. To solve this problem, the authors introduce pixel-level histogram loss to constrain the similarity. The perceptual loss and cycle consistency loss are used to constrain the preserve identity. Finally, the authors also propose a new dataset, the MT dataset.



**FIGURE 5** Framework diagram of the makeup transfer (MT) method inspired by CycleGAN. BeautyGAN achieves automatic MT with a dual input/output Generative Adversarial Network. In addition, it also achieves pleasant makeup images and accurate results. PairedCycleGAN [71] is an unsupervised method. It uses the same framework for training both MT and removal networks

### 2.3.2 | PairedCycleGAN

Inspired by CycleGAN, the authors of Ref. [71] propose a cycle-consistent GAN by proposing two asymmetric functions. One is responsible for makeup and the other for de-makeup. It is an unsupervised method that eliminates a large number of preprocessing steps compared to previous methods based on image processing. PairedcycleGAN trains three generators and discriminators for each facial component. In addition, the proposed de-makeup network F can remove detected cosmetics to recover the original face. The main uses are adversarial loss, identity loss, and style loss. The authors conclude by proposing further applications of the PairedCycleGAN model, such as automatic ageing, automatic de-ageing etc.

## 2.4 | Makeup transfer method optimised for high-quality images

This type of method can achieve a good effect of MT and can output a high quality image in the end. However, the robustness of these methods is poor, and the effect of MT is poor when the facial pose and expression changes are large. The network structure of DMT [73] and RAMT-GAN [74] is shown in Figure 6.

### 2.4.1 | DMT

Inspired by disentangled representation, the authors of Ref. [73] propose a GAN DMT for MT in different scenarios. The model

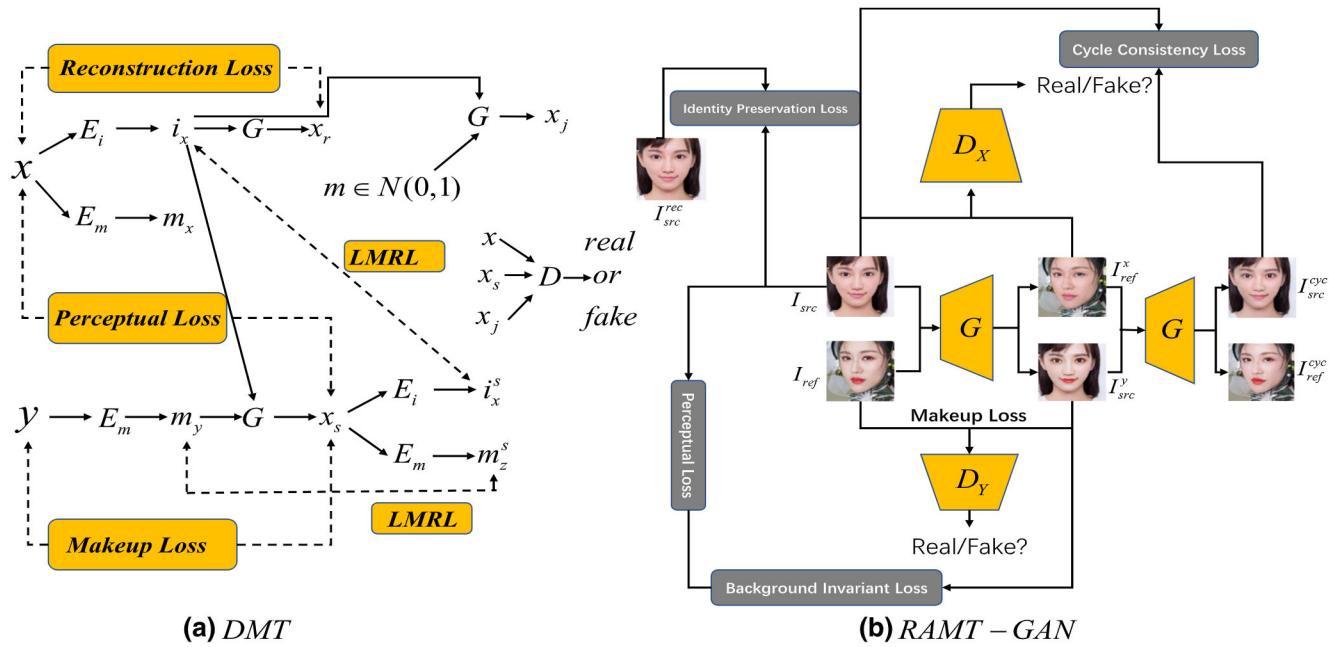
includes an identity encoder and a makeup encoder, and a decoder is used to reconstruct the face based on the output of both encoders. In addition, a distinguisher is used to identify the faces to achieve a high-quality face output. It is worth noting that the framework of DMT, which can implement different MT scenarios, includes pairing, interpolation, blending etc.

### 2.4.2 | RAMT-GAN

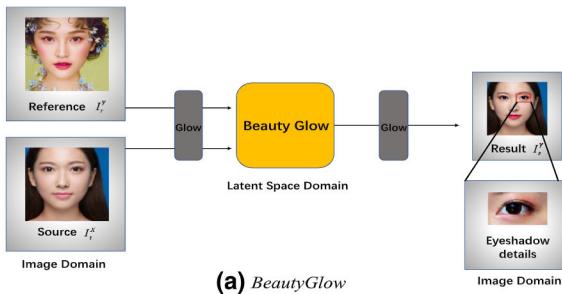
Most of the current makeup transfer methods lack the generated face images with poor quality. The authors of Ref. [74] proposed a GAN-based makeup transfer method, called RAMT-GAN. This is an unsupervised method to achieve realistic and accurate MT. While completing MT, background information and face identity are preserved. To address identity-shift and background-change, the authors introduce two loss functions. Finally, it is experimentally shown that the proposed method outperforms several existing methods in terms of image quality.

## 2.5 | Other makeup transfer methods

In addition to the four types of methods introduced above, there are some other methods that can also achieve the effect of MT. For example, SCGAN, which can achieve accurate MT, and BeautyGlow, which is the first glow-based MT method, can finally generate high-quality makeup images. The network framework diagram of BeautyGlow is shown in Figure 7.



**FIGURE 6** Framework diagram of the makeup transfer (MT) method optimised for high-quality images. DMT is the first method to integrate disentangled representation to complete the facial MT. RAMT-GAN is also an unsupervised method that can achieve accurate MT. As the figure shows, it introduces two loss functions in the architecture



**FIGURE 7** Framework diagram of the other makeup transfer (MT) methods. BeautyGlow is the first method to complete MT-based glow. Additionally, through manipulating on latent vectors, it can generate realistic images from light makeup to heavy makeup

### 2.5.1 | SCGAN

Before SCGAN was proposed, most methods implemented MT by computing the pixel level between two images. However, the computational cost of this method is high. To address this problem, the authors of Ref. [75] proposed a style-controlled GAN model. This model consists of three parts, each for target style-code encoding, face identity features extraction, and makeup fusion, respectively. SCGAN is a more flexible model for automatic MT. In addition, to eliminate the problem of spatial misalignment, a new design is proposed. First, the MT problem is divided into two steps of extraction and assignment. In addition, a style-based network PSEnc is proposed, which can map makeup style to a component-wise style-code. In general, the method can achieve good effect of MT by editing the style-code but performs generally for large expressions and large poses.

### 2.5.2 | BeautyGlow

Due to the huge number of makeup combinations, in order to achieve on-demand makeup, the authors of Ref. [76] proposed BeautyGlow. BeautyGlow is the first Glow-based MT model. Inspired by Glow, the authors help training by building new transfer matrices and loss functions. BeautyGlow can decompose the latent vectors of face images derived from the Glow model into makeup and non-makeup latent vectors. Makeup and de-makeup are done by manipulating the latent space. In order to obtain better results for the MT, the authors introduce a series of loss functions to guide the function decomposition to help generate high-quality MT images. Notably, the authors use perceptual loss. What's more, intra-domain loss and inter-domain losses are designed for ensuring the after-makeup image and de-makeup image. Makeup loss is used for extracting the features of makeup. Circular consistency loss is used to maintain facial and makeup information.

## 3 | INTRODUCTION TO THE LOSS FUNCTION

The loss function plays a key role in the quality of the final generated image. Not only can the loss function be used to measure the difference between the synthetic portrait image and the reference makeup but it can also be used to constrain the identity and makeup information. Through systematic literature research, the commonly used loss functions are summarised in this paper. They mainly include makeup loss, adversarial loss, and cyclic consistency loss. To facilitate the description,  $X$  and  $Y$  are used in this section to denote the

source image domain and the reference image domain, respectively,  $c$  refers to the target colour of the makeup style,  $G$  represents the generator, and  $D$  represents the discriminator.

### 3.1 | Adversarial loss

In order to improve the quality of the final image generation, based on the principle of fighting against loss, the authors introduce a generator and a discriminator. These two, in a continuous game, are continuously trained, and as the number of iterations increases, the generated image of the generator gets closer and closer to the real image, thus achieving the effect of improving the quality of the image. In any GAN problem, the authors use adversarial loss, whose goal is to make the generated image indistinguishable from the real image. For the discriminator, the authors use the adversarial loss described in Equation (1), and for the generator, the authors use the adversarial loss described in Equation (2):

$$L_a^D = \frac{1}{n} \sum_{i=1}^n D_p(G(x_i, c_i)) - \frac{1}{n} \sum_{i=1}^n D_p(x_i) + \lambda_{gp} gp(D) \quad (1)$$

$$L_a^G = -\frac{1}{n} \sum_{i=1}^n D_p(G(x_i, c_i)) \quad (2)$$

where  $D_p$  is the positive classification output of the discriminator and  $\lambda_{gp} gp(D)$  is the weighted gradient penalty term computed on  $D$ .

### 3.2 | Cycle consistency loss

The acquisition of the triplet data (source image, reference image and transmission image) is difficult. Therefore, the network is usually trained in an unsupervised manner. In order to maintain the facial and makeup information, two cycles of consistency loss are also designed in the latent space. Specifically, if the authors multiply the latent vector of the make-up, that is,  $I_s^Y$ , with the transfer matrix  $W$ , it should be close to the facial latent vector of the source image, that is,  $F_s^Y$ . On the other hand, if the authors multiply with  $(I - W)$ , it is considered to be close to the constitutive latent feature of the reference latent feature. Thus, the loss function can be written as

$$\mathcal{L}_{cyc} = \|L_s^Y W - F_s^X\|_2 + \|L_s^Y (I - W) - M_r^Y\|_2 \quad (3)$$

### 3.3 | Perceptual loss

When performing the MT task, the identity information of the makeup images needs to be preserved. Therefore, the identity information of the source images needs to be guaranteed by introducing a perceptual function. In contrast to the direct

calculation of the difference between the pixel-level Euclidean distance, the perceptual loss is used to compare the activation of the source image and the image generated by the hidden layer using a VGG – 16 [77] model pre-trained on ImageNet. Let be the output of the VGG – 16 model at layer  $l$  and introduce perceptual loss to measure the difference in loss:

$$\begin{aligned} L_G^{per} = & \mathbb{E}_{x \sim \mathcal{P}_X, y \sim \mathcal{P}_Y} [\|F_l(G(x, y)) - F_l(x)\|_2] \\ & + \mathbb{E}_{x \sim \mathcal{P}_X, y \sim \mathcal{P}_Y} [\|F_l(G(y, x)) - F_l(y)\|_2] \end{aligned} \quad (4)$$

where  $x$  is the sample selected from  $x$  according to the distribution  $P(x)$  and  $y$  is the sample selected from  $y$  according to the distribution  $P(y)$ .

### 3.4 | Makeup loss

The overall makeup loss consists of four local histogram losses acting on the face, lips, eyes and eyebrow areas.

$$L_m = \lambda_{face} L_{face} + \lambda_{lip} L_{lip} + \lambda_{eye} L_{eye} + \lambda_{brow} L_{brow} \quad (5)$$

where  $\lambda_{face}$ ,  $\lambda_{lip}$ ,  $\lambda_{eye}$  and  $\lambda_{brow}$  are weighting factors. To further encourage the network to convey the makeup components of each face, the constraint of makeup style consistency should be considered in the network. Spatial histogram matching of the reference image and the generated image is performed by multiplying the image with the corresponding binary mask, and a face resolution model is used to obtain the face bootstrap mask  $M_{item} = FP(x)$ , where  $item \in \{ \text{lips, eye, face, brow} \}$ . Thus, the local histogram loss is defined as

$$\begin{aligned} L_{item} &= \|I_{src}^y - HM(I_{src} \circ M_{item}^1, I_{ref} \circ M_{item}^2)\|_2 \\ M^1 &= FP(I_{src}), M^2 = FP(I_{ref}) \end{aligned} \quad (6)$$

### 3.5 | Other loss

#### 3.5.1 | Identity loss

Identity loss is introduced in order to maintain the identity information of the source image during the MT. The principle of identity loss is similar to the cyclic consistency loss introduced earlier. The identity-preserving loss function is roughly defined to minimise the identity gap between the input image and the reconstructed image, thus preserving the identity of the face. First, a face resolution mask needs to be used. For each input image  $x$ , the trained face resolution model generates an index template  $M$  that represents several facial locations, including lips, eyes, eyebrows, facial skin, background etc. The original foreground mask  $M_{fore}$  is multiplied with  $I_{src}$ ,  $I_{src}^{rec}$  to obtain two foreground images  $f_{src}, f_{src}^{rec}$ , and the absolute difference is calculated using  $L1$  parametric loss:

$$L_{fore\_src} = \|f_{src}^{rec} - f_{src}\|_1 \quad (7)$$

### 3.5.2 | Style loss

Style losses are used to ensure the successful migration of the details of a specific makeup style. Since the previous losses limit the output of  $G$ , they are not sufficient to ensure successful transfer of the details of a specific makeup style  $Y^\beta$ . For this reason, PairedCycleGAN proposes two style losses,  $L_1$  reconstruction loss  $L_S$  and style discrimination loss  $L_P$ . First  $L_S$  is used to ensure that the closer the reconstructed image of the reference image to the reference image, the better the formula

$$L_S = \mathbb{E}_{x \sim \mathcal{P}_X, y^\beta \sim \mathcal{P}_Y} [\|G(F(y^\beta), G(x, y^\beta)) - y^\beta\|_1] \quad (8)$$

Using  $L_1$  loss in the pixel domain can help convey the overall structure and colour (e.g. the shape of the eyebrows and the gradient of the eye shadow) but can lead to blurred results that do not convey sharp edges (e.g. eyelashes and eyeliner). Therefore, it is necessary to add an auxiliary discriminator  $D_S$ , which determines whether a given pair of faces is made up or not. The loss function of the auxiliary discriminator  $D_S$  is

$$\begin{aligned} L_P = & \mathbb{E}_{x \sim \mathcal{P}_X, y^\beta \sim \mathcal{P}_Y} [\log D_S(y^\beta, W(x, y^\beta))] \\ & + \mathbb{E}_{x \sim \mathcal{P}_X, y^\beta \sim \mathcal{P}_Y} [\log(1 - D_S(y^\beta, G(x, y^\beta)))] \end{aligned} \quad (9)$$

### 3.5.3 | Asymmetric losses

While lighter styles of MT tasks mainly involve the recolouring of eyeshadows and lips, extreme makeup styles present new challenges in this issue. On the one hand, extreme makeup styles contain high-frequency components and the network needs to distinguish other high-frequency facial textures (e.g. eyelashes). On the other hand, in some extreme de-makeup cases, it is difficult to observe the original facial colour of a person in the image after de-makeup, which requires the network to reconstruct the facial skin colour without makeup. To address these challenges, the authors add a higher-order loss  $L^{ho}$  in the MT branch to help transfer high-frequency details and a smoothing loss  $L^{smooth}$  in the de-makeup branch, which is based on the assumption that facial colours after de-makeup are usually smooth.

**Higher-order loss:** Since the distorted image  $W(x_i, y_j)$  retains most of the texture information of the make-up style (colour change and edges) of the reference image  $y_j$ , the authors calculate the higher order loss by applying a Laplace filter to the local block, which is calculated as

$$L^{ho} = \sum_k h_k \|f(p_k^W) - f(\tilde{p}_k^Y)\|_1 \quad (10)$$

where  $h_k$  is the weight of the local block and  $f$  is the Laplacian filter.

### Smoothing loss

In contrast to the make-up transfer result ( $y_j$ ), the authors do not want the unmakeup result ( $x_i$ ) to have high-frequency details. Instead, it should be smooth in the local part. Therefore, the authors apply a smoothing loss to  $x_i$ , which is calculated as

$$L^{smooth} = \sum_k s_k \|f(\tilde{p}_k^X)\|_1 \quad (11)$$

where  $\tilde{p}_k^X$  is the local block of the de-makeup image,  $s_k$  is the weight of this local block, and  $f$  is the Laplacian filter. Unlike the higher order losses, the weights for the eye region need to be smaller to preserve the high-frequency texture around the eyes. For the cheek and nose regions, larger weights need to be given so that these regions are smoother. The smoothing loss constrains the high-frequency texture to appear in  $x_i$ , while the higher-order loss tries to merge that weight into  $y_j$ . By introducing these asymmetric losses, the authors are eventually able to generate makeup images with nice effects.

### 3.5.4 | Loss of makeup details

In the MT task, makeup loss, while providing constraints at the face region level, this absence makes it difficult to migrate some makeup details, such as highlight and blush. To address this problem, PSGAN++ proposes a novel makeup detail loss. First, the authors apply the dense facial alignment method to detect the dense facial feature points of source image  $x$  and reference image  $y_1$ . Then, the authors select K feature points located in the makeup detail region, that is, nose and cheeks, and compose them into makeup detail markers  $L_x$  and  $L_{y1}$ . Finally, the makeup detail loss is calculated by the selected K feature points. The makeup detail loss is calculated as the difference between the corresponding makeup detail feature points between the MT image  $T(x, y_1)$  and the reference image  $y_1$ , which is calculated as

$$L_G^{\text{detail}} = \mathbb{E}_{x \sim \mathcal{P}_X, y_1 \sim \mathcal{P}_Y} \left[ \sum_k \|T(x, y_1)^k - y_1^k\|_1 \right] \quad (12)$$

where  $T(x, y_1)^k$  and  $y_1^k$  denote the pixel values of the corresponding Kth feature point in  $T(x, y_1)^k$  and  $y_1^k$ , respectively.

### 3.5.5 | Histogram matching loss

The purpose of histogram matching loss is to match the colour distribution of the reference image  $y$  after MT with the source image  $x$ . It is the key loss function proposed by BeautyGAN for migrating makeup colours in the makeup region. Directly applying the pixel-level histogram of both images with the mean square error loss does not achieve the optimisation purpose, because the gradient can be obtained as 0 according to the relationship of the indicator function. Thus, the authors use the



histogram matching strategy to change the histogram of the source image by using the histogram matching function for matching the reference image in the pre-defined regions, including facial skin, eye shadow and lips, where the total loss is the weighted sum of the losses in each region and is calculated as

$$\mathcal{L}_{\text{hist}} = \lambda^{\text{eyes}} \mathcal{L}_{\text{hist}}^{\text{eyes}} + \lambda^{\text{lips}} \mathcal{L}_{\text{hist}}^{\text{textlips}} + \lambda^{\text{skin}} \mathcal{L}_{\text{hist}}^{\text{skin}} \quad (13)$$

where  $\lambda^{\text{skin}}$ ,  $\lambda^{\text{eyes}}$ , and  $\lambda^{\text{lips}}$  are adjustable hyper parameters, respectively. The loss represented by each of these parameters is the difference between the makeup-migrated reference image and the histogram-matched image, which is calculated as

$$\mathcal{L}_{\text{hist}}^i = \|T_s^{mC} \odot \Gamma_s^i - HM(T_s^n \odot \Gamma_s^i, T_r^n \odot \Gamma_r^i)\| \quad (14)$$

where the symbol  $\odot$  denotes the pixel-by-pixel multiplication, and  $\Gamma_s^i$  and  $\Gamma_r^i$  denote the segmentation masks of  $i$  regions in the source and reference images, respectively.

## 4 | INTRODUCTION OF THE DATASET

The rapid growth in the field of MT has been made possible by the advent of generative adversarial networks and a large number of high-quality datasets. The availability of high-quality datasets has allowed research to advance rapidly. As new methods have been proposed, the datasets have changed considerably. The datasets can be broadly divided into two types, one with paired pre-makeup and makeup and the other with unpaired datasets. With the development in the field, the methods based on image processing and supervised data requiring paired data are gradually replaced by GAN-based methods. At the same time, a large number of new datasets have emerged, with three differences relative to the previous ones. First, a larger number of large models with better results emerge, but a larger number of datasets are needed for training; second, higher quality images; and third, higher resolution images. We list the commonly used MT datasets in Table 1.

### 4.1 | Makeup transfer

MT was proposed along with BeautyGAN back in 2018. The dataset included a total of 3834 images of women. Among them, 1115 non-makeup images and 2719 makeup images were included. The transfers of background, race, pose, and expression are included. Also, there are more makeup styles, such as retro makeup, Japanese costume, and Korean costume. In addition, there is also diversity in the degree of makeup, with heavier makeup and lighter makeup. The data is collected from the Internet, and finally, the images with low quality and low resolution are removed by human culling. Moreover, the position of the images was adjusted using key point detection technology. Compared with other datasets containing before and after makeup images, MT is more suitable for performing MT tasks.

**TABLE 1** Commonly used makeup transfer datasets in makeup transfer

Number	Dataset	Time	Paired images	Number
1	YMU	2012	Yes	604
2	VMU	2012	Yes	204
3	Makeup transfer(MT)	2018	No	3834
4	LADN	2019	No	635
5	Makeup – Wild	2020	No	772
6	CPM	2021	No	11,652

### 4.2 | LADN makeup

Previously face image datasets are usually more suitable for recognition or identity verification tasks, and only a few makeup datasets are publicly available and have low resolution or image quality. To solve this problem, the authors propose the LADN Makeup dataset. The generation process of this makeup dataset is roughly given as follows: first, unobstructed facial images are collected from the Internet. Then, non-frontal faces are filtered out, and finally 333 pre-makeup images and 302 post-makeup images are included. In order to obtain some extreme makeup looks, 115 additional extreme style makeup images were added manually.

### 4.3 | Makeup-Wild

Makeup-Wild was introduced in 2020, along with PSGAN. Prior to that, most of the makeup datasets contained only positive data of neutral expressions. To solve this problem, the authors first collected a large number of face images from the Internet, after which the images with positive or neutral expressions were manually removed. Then, the size of the images was cropped to a resolution of  $256 \times 256$  size. Finally, 403 makeup images and 369 non-makeup images containing various poses and expressions and complex backgrounds of faces were obtained.

### 4.4 | CPM

CPM is a dataset with diversity; it contains four sub-datasets called CPM-Real, CPM-Synt-1, CPM-Synt-2 and Stickers. CPM-Real is a dataset with real faces and with extreme styles of makeup. It has very diverse makeup styles and contains both colour and pattern makeup styles. To obtain this dataset, the authors first obtained the initial images from the web using keywords, then detected and cropped the faces using the MTCNN [78] face detector, discarding face images with a resolution of less than  $150 \times 150$ . Then, the low-quality face images were manually removed, and finally a CPM-Real makeup dataset containing 3895 face images was obtained, which was mainly used for conducting tests. CPM-Synt-1 is a synthetic face image dataset. First, some patterns were

collected from the web as Stickers dataset. The Stickers dataset contains 577 high quality images. After that, these patterns are suitably applied to the MT dataset to finally get 5555 images after makeup. The CPM-Synt-2 dataset is mainly a synthetic dataset for the evaluation of makeup effect. The CPM-Synt-2 consists of 1625 ternary images, and each set of images contains the source image, the reference image, and the pattern corresponding to the ground-truth.

#### 4.5 | Example of dataset partitioning for makeup transfer

To help the less experienced reader understand the use of datasets in MT tasks, we will present an example of an MT task. Most of the current datasets in the field of MT are unsupervised datasets, which are further subdivided into datasets that have paired face images or not. For example, YMU and VMU are both images with before-and-after makeup pairing, which are from the same face, while the surrounding conditions are relatively constant to ensure that fewer variables are changed. However, these MT methods that require before-and-after paired images are currently obsolete due to too many limitations, and the mainstream methods are now based on GAN implementations. Training is usually performed on datasets that do not contain paired face images. For example, training is performed on the CPM-Real dataset, which contains 3895 images with makeup on. It is noteworthy to expand on the MT domain, for example, to migrate the face makeup style while transforming the stickers and patterns at the same time. The implementation of this process is done with the help of supervised datasets, for example, CPM introduces both CPM-Synt-1 and CPM-Synt-2, and Stickers datasets in order to achieve the transfer of stickers and patterns. As an example, the division of the dataset of PSGAN++ uses face images of image size  $512 \times 512$  as the dataset, of which 2000 are makeup images and 1000 are non-makeup images. The authors randomly selected 10% (200) of the makeup images and 10% (100) of the non-makeup images from them as test data, and the rest of the images were used for training and validation. For example, RAMT used a  $361 \times 361$  image size as the data set, which contained 1115 non-makeup images and 2719 makeup images, with a full coverage of makeup styles, including natural and make-up, Korean makeup, and Japanese makeup. Finally, 100 of these non-makeup images and 250 makeup images were used as the test set, and the rest were used as training and validation. Thus, in general, 10% of the total image data is used for testing, and the remaining 90% is generally used for training and validation.

### 5 | EVALUATION INDICATORS

At present, the evaluation of the effect of MT is mainly of two types. The first type is qualitative evaluation, that is, the effect of MT is judged by manual means. The second category is quantitative evaluation, which is more objective.

#### 5.1 | Qualitative evaluation

The effect of the MT is evaluated directly by means of the human eye. This way of evaluation is more subjective, mainly by observing whether the makeup is natural as a whole and whether the local details are properly handled. Comparison of details is made by the naked eye.

#### 5.2 | Quantitative evaluation

User style means that several researchers are selected to judge the final generated facial makeup images. This method generally considered the quality, realism, and similarity of makeup styles. This evaluation method is the most common in the field of MT.

### 6 | PROSPECT

After a systematic review of the field, this paper considers the following possible directions for the future development of the MT field.

#### 6.1 | Multi-view consistent 3D face makeup transfer

Currently, MT methods end up generating 2-dimensional makeup application images. However, in reality, the evaluation of makeup effect should be multi-view, and a good makeup should have a good effect in several views. Therefore, the evaluation of the MT effect should also be multi-view, and currently, most methods apply makeup to the frontal face picture. In the field of MT, from the very beginning of simple pixel editing (dots) to the modification of local makeup effects (lines) and to the combination of subsequent makeup effects (faces), the trend of future development should be from the current 2-dimensional MT effect to the future 3-dimensional MT.

#### 6.2 | Building a high-quality makeup transfer dataset

The rapid development of a field is inseparable from excellent datasets. The development of datasets will largely promote the emergence of many new methods. For example, the previous generation of MT methods are mostly based on graphics processing and makeup paired datasets. In recent years, with the proposal of GAN and the emergence of a large number of unsupervised datasets, most of the current methods are implemented based on GAN and do not require paired datasets before and after makeup. However, most of the current makeup datasets have certain problems, such as insufficient quantity, low overall quality, low resolution etc. With the continuous proposal of various models, many large models

have good results, but the training of large models requires enough high-resolution facial makeup images. Therefore, the construction of high-quality datasets plays a critical role in the development of the entire field. Constructing high-quality MT datasets will be a research hotspot in the future, and with the emergence of high-quality datasets, a series of new methods will be generated to deal with MT tasks.

### 6.3 | A way to change makeup with controlled intensity

At present, makeup conversion is mainly based on GAN, which is trained to generate high-quality makeup images. But the makeup intensity of the resulting image cannot be precisely controlled. The intensity of MT should be intensity controllable so as to ensure that different intensities of makeup are generated, which will provide more possibilities for the implementation of MT methods. The stable and controllable method provides technical support for the subsequent interactive application of MT, helping users to customise and obtain different MT strengths to form more makeup styles. Due to its unique training method, the GAN network is difficult to converge, and it is difficult to achieve stability and control. At present, there are also some attempts, such as introducing supervised datasets to help it converge, such as loading pre-training models and combining multiple tasks. Way. A stable and controllable MT method has great application value and is the focus of the next generation of MT methods.

### 6.4 | Multi-makeup makeup transfer method

The so-called multi-makeup MT method refers to the ability to generate makeup that did not exist before on the basis of the original so as to further improve the diversity of makeup styles. For example, the new makeup is composed of the face colour of makeup A, the lip colour of makeup B, and the eye shadow style of makeup C. That is, it can be understood as the generation of a new makeup style. Currently, most methods in the field of MT can only perform MT on one reference image and one source image. If the MT of multiple reference images and one source image can be realised, makeup that is not originally available will be generated, which can further shorten the makeup required for trial. The task of makeup style generation has not received the attention of scholars, but this task has strong practicability and high landing value and can expand the original makeup style.

### 6.5 | Makeup recommendation system

At present, there are a large number of different types of makeup styles, but for different people, the same person, different occasions, and even people's moods, it is necessary to match the appropriate makeup to maximise the attractiveness

of the appearance. In fact, due to the variety of makeup styles, choosing the right makeup is not an easy task. Second, because makeup looks different on different people's faces, it is difficult to make up the style of makeup through personal imagination. At present, most of the MT methods are based on GAN, and the methods are relatively mature. Some makeup transfer methods have strong robustness and can achieve good results under complex poses and expressions. Works well even in the presence of shadows and occlusions. This paved the way for the development of the follow-up MT recommendation system. At present, the makeup change recommendation system is not yet mature, and it is difficult to give professional makeup suggestions. In previous years, there was a lack of makeup image datasets. In recent years, a large number of MT image datasets have been proposed and widely used. Therefore, MT recommendation system is an important research topic. With the development of modern aesthetics, more and more makeup looks appear, which makes people dazzling, and it is impossible to quickly choose the makeup that suits them. If it can be recommended by the machine, the practicality of makeup will be further improved.

## 7 | SUMMARY

Since ancient times, people have had a keen interest in the beauty of the face. Makeup is an important means of enhancing facial beauty, and limited by various resources, people cannot exhaust all makeup styles. With the continuous development of makeup changing technology, it has made virtual fitting a reality. Therefore, MT technology has significant theoretical research value and commercial landing applications. We summarise the classic makeup methods for the first time. This paper provides a comprehensive survey and discussion of MT in recent years. First, we introduce the definition of MT. Then, we discuss about the classic MT methods. Second, the evaluation methods are introduced. The evaluation of the effect of MT is mainly of two types. And then, we compare the existing methods, and existing datasets are introduced. Finally, we discuss the current problems of previous makeup methods and the possible future directions.

## ACKNOWLEDGEMENTS

This work was supported in part by the Scientific Research Projects of the Hubei Provincial Department of Education (D20201304), Open Fund of Xi'an Key Laboratory of Tight oil (Shale oil) Development XSTS-202101, the Open Foundation of Cooperative Innovation Center of Unconventional Oil and Gas, Yangtze University (Ministry of Education & Hubei Province), No. UOG2022-06, and the National Natural Science Foundation of China under Grant No. 51974036 and No.62006028.

## CONFLICT OF INTEREST

The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.



## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available, reference number [62].

## ORCID

Feng He  <https://orcid.org/0000-0001-5059-4423>

## REFERENCES

1. Abadi, M., et al.: {TensorFlow}: a system for {Large-Scale} machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp. 265–283 (2016)
2. Ning, X., et al.: Feature refinement and filter network for person re-identification. *IEEE Trans. Circ. Syst. Video Technol.* 31(9), 3391–3402 (2020). <https://doi.org/10.1109/tcsvt.2020.3043026>
3. Xiao, B., et al.: Explainable deep learning for efficient and robust pattern recognition: a survey of recent developments. *Pattern Recogn.* 120, 108102 (2021). <https://doi.org/10.1016/j.patcog.2021.108102>
4. Cheng, K., et al.: Logical topology inference via CPGCN joint optimizing with pedestrian re-ID. *IEEE Transact. Neural Networks Learn. Syst.*, 1–13 (2021). <https://doi.org/10.1109/tnnls.2021.3125368>
5. Paulo Lima, J., et al.: Generalizable multi-camera 3D pedestrian detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1232–1240 (2021)
6. Ning, X., et al.: JWSAA: joint weak saliency and attention aware for person re-identification. *Neurocomputing* 453, 801–811 (2021). <https://doi.org/10.1016/j.neucom.2020.05.106>
7. Zhai, W., Peng, Z.R.: Damage assessment using google street view: evidence from hurricane Michael in Mexico beach, Florida. *Appl. Geogr.* 123, 102252 (2020). <https://doi.org/10.1016/j.apgeog.2020.102252>
8. Meng, Q., et al.: Magface: a universal representation for face recognition and quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14225–14234 (2021)
9. Zheng, Z., et al.: Webface260m: a benchmark unveiling the power of million-scale deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10492–10502 (2021)
10. Wang, C., et al.: Uncertainty estimation for stereo matching based on evidential deep learning. *Pattern Recogn.* 124, 108498 (2022). <https://doi.org/10.1016/j.patcog.2021.108498>
11. Ning, X., et al.: Real-time 3D face alignment using an encoder-decoder network with an efficient deconvolution layer. *IEEE Signal Process. Lett.* 27, 1944–1948 (2020). <https://doi.org/10.1109/lsp.2020.3032277>
12. He, F., Ye, Q.: A bearing fault diagnosis method based on wavelet packet transform and convolutional neural network optimized by simulated annealing algorithm. *Sensors* 22(4), 1410 (2022). <https://doi.org/10.3390/s22041410>
13. Ding, Y., et al.: Semi-supervised locality preserving dense graph neural network with ARMA filters and context-aware learning for hyperspectral image classification. *IEEE Trans. Geosci. Rem. Sens.* 60, 1–12 (2021). <https://doi.org/10.1109/tgrs.2021.3100578>
14. Ding, Y., et al.: Graph sample and aggregate-attention network for hyperspectral image classification. *IEEE Geosci. Rem. Sens. Lett.* 19, 1–5 (2021). <https://doi.org/10.1109/lgrs.2021.3062944>
15. Yao, Y., et al.: Exploiting web images for dataset construction: a domain robust approach. *IEEE Trans. Multimed.* 19(8), 1771–1784 (2017). <https://doi.org/10.1109/tmm.2017.2684626>
16. Yao, Y., et al.: Towards automatic construction of diverse, high-quality image datasets. *IEEE Trans. Knowl. Data Eng.* 32(6), 1199–1211 (2019). <https://doi.org/10.1109/tkde.2019.2903036>
17. Yao, Y., et al.: Non-salient region object mining for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2623–2632 (2021)
18. Wang, Z., et al.: GLStyleNet: exquisite style transfer combining global and local pyramid features. *IET Comput. Vis.* 14(8), 575–586 (2020). <https://doi.org/10.1049/iet-cvi.2019.0844>
19. Chen, H., et al.: Creative and diverse artwork generation using adversarial networks. *IET Comput. Vis.* 14(8), 650–657 (2020). <https://doi.org/10.1049/iet-cvi.2020.0014>
20. Xu, Z., et al.: Stroke controllable style transfer based on dilated convolutions. *IET Comput. Vis.* 14(7), 505–516 (2020). <https://doi.org/10.1049/iet-cvi.2019.0912>
21. Horita, D., Aizawa, K.: SLGAN: style-and latent-guided generative adversarial network for desirable makeup transfer and removal. *arXiv preprint arXiv:2009.07557*, (2020)
22. Wei, X.S., et al.: Fine-Grained Image Analysis with Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021)
23. Ning, X., et al.: HCFNN: High-Order Coverage Function Neural Network for Image Classification. *Pattern Recognition*. 108873 (2022)
24. Cai, C., et al.: A cost-effective segmented dynamic wireless charging system with stable efficiency and output power. *IEEE Trans. Power Electron.* 37(7), 8682–8700 (2022). <https://doi.org/10.1109/tpe.2022.3143128>
25. Cai, C., et al.: A multichannel wireless UAV charging system with compact receivers for improving transmission stability and capacity. *IEEE Syst. J.* 16(1), 997–1008 (2021). <https://doi.org/10.1109/jsyst.2021.3085914>
26. Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, pp. 341–346 (2001)
27. Drobi, I., Cohen-Or, D., Yeshurun, H.: Example-based style synthesis. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, vol. 2, pp. II–143. IEEE (2003)
28. Frigo, O., et al.: Split and match: example-based adaptive patch sampling for unsupervised style transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 553–561 (2016)
29. Elad, M., Milanfar, P.: Style transfer via texture synthesis. *IEEE Trans. Image Process.* 26(5), 2338–2351 (2017). <https://doi.org/10.1109/tip.2017.2678168>
30. Gooch, B., Gooch, A.: Non-photorealistic Rendering. AK Peters/CRC Press (2001)
31. Strothotte, T., Schlechtweg, S.: Non-photorealistic Computer Graphics: Modeling, Rendering, and Animation. Morgan Kaufmann (2002)
32. Paul, R., Collomosse, J.: Image and Video-Based Artistic Stylisation, vol. 42. Springer Science & Business Media (2012)
33. Zhao, H.H., et al.: Automatic semantic style transfer using deep convolutional neural networks and soft masks. *Vis. Comput.* 36(7), 1307–1324 (2020). <https://doi.org/10.1007/s00371-019-01726-2>
34. Li, R., et al.: SDP-GAN: saliency detail preservation generative adversarial networks for high perceptual quality style transfer. *IEEE Trans. Image Process.* 30, 374–385 (2020). <https://doi.org/10.1109/tip.2020.3036754>
35. Chen, X., et al.: Infogan: interpretable representation learning by information maximizing generative adversarial nets. *Adv. Neural Inf. Process. Syst.* 29 (2016)
36. Nguyen, T.V., Liu, L.: Smart mirror: intelligent makeup recommendation and synthesis. In: Proceedings of the 25th ACM International Conference on Multimedia, pp. 1253–1254 (2017)
37. Alashkar, T., et al.: Examples-rules guided deep neural network for makeup recommendation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31 (2017)
38. Alashkar, T., Jiang, S., Fu, Y.: Rule-based facial makeup recommendation system. In: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 325–330. IEEE (2017)
39. Chung, K.Y.: Effect of facial makeup style recommendation on visual sensibility. *Multimed. Tool. Appl.* 71(2), 843–853 (2014). <https://doi.org/10.1007/s11042-013-1355-6>
40. Liu, L., et al.: Wow! you are so beautiful today. *ACM Trans. Multimed. Comput. Commun. Appl.* 11(1s), 1–22 (2014). <https://doi.org/10.1145/2659234>
41. Liao, J., et al.: Visual attribute transfer through deep image analogy. *arXiv preprint arXiv:1705.01088*, (2017)

42. Luan, F., et al.: Deep photo style transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4990–4998 (2017)
43. Taigman, Y., Adam, P., Wolf, L.: Unsupervised cross-domain image generation. arXiv preprint arXiv:1611.02200, (2016)
44. Chen, C., Dantcheva, A., Ross, A.: Automatic facial makeup detection with application in face recognition. In: 2013 International Conference on Biometrics (ICB), pp. 1–8. IEEE (2013)
45. Chen, C., Dantcheva, A., Ross, A.: An ensemble of patch-based subspaces for makeup-robust face recognition. Inf. Fusion 32, 80–92 (2016). <https://doi.org/10.1016/j.inffus.2015.09.005>
46. Chen, C., et al.: Spoofing faces using makeup: an investigative study. In: 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), pp. 1–8. IEEE (2017)
47. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, pp. 187–194 (1999)
48. Ojima, N., et al.: Image synthesis of cosmetic applied skin based on optical properties of foundation layers. In: Proceedings of International Congress of Imaging Science, pp. 467–468 (1999)
49. Leyvand, T., et al.: Digital face beautification. In: ACM Siggraph 2006 Sketches. 169–es (2006)
50. Guo, D., Sim, T.: Digital face makeup by example. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 73–79. IEEE (2009)
51. Scherbaum, K., et al.: Computer-suggested facial makeup. In: Computer Graphics Forum, vol. 30, pp. 485–492. Wiley Online Library (2011)
52. Chen, L., Zhou, K., Lin, S.: Simulating makeup through physics-based manipulation of intrinsic image layers. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4621–4629 (2015)
53. Choi, Y., et al.: Stargan: unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8789–8797 (2018)
54. Dantcheva, A., Chen, C., Ross, A.: Can facial cosmetics affect the matching accuracy of face recognition systems? In: 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 391–398. IEEE (2012)
55. Johnson, J., Alexandre, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision, pp. 694–711. Springer (2016)
56. Liu, S., et al.: Makeup like a superstar: deep localized makeup transfer network. arXiv preprint arXiv:1604.07102, (2016)
57. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39(4), 640–651 (2016). <https://doi.org/10.1109/tpami.2016.2572683>
58. Chen, Y.-C., Shen, X., Jia, J.: Makeup-go: blind reversion of portrait edit. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4501–4509 (2017)
59. Gatys, L.A., et al.: Preserving color in neural artistic style transfer. arXiv preprint arXiv:1606.05897, (2016)
60. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576, (2015)
61. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414–2423 (2016)
62. Goodfellow, I., et al.: Generative adversarial nets. Adv. Neural Inf. Process. Syst. 27 (2014)
63. Tewari, A., et al.: Mofa: model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 1274–1283 (2017)
64. Jiang, W., et al.: Psgan: pose and expression robust spatial-aware gan for customizable makeup transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5194–5202 (2020)
65. Gu, Q., et al.: Ladn: local adversarial disentangling network for facial makeup and de-makeup. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10481–10490 (2019)
66. Lyu, Y., et al.: SOGAN: 3D-aware shadow and occlusion robust GAN for makeup transfer. In: Proceedings of the 29th ACM International Conference on Multimedia, pp. 3601–3609 (2021)
67. Liu, S., et al.: Psgan++: Robust Detail-Preserving Makeup Transfer and Removal. IEEE Transactions on Pattern Analysis and Machine Intelligence (2021)
68. Nguyen, T., Tran, A.T., Hoai, M.: Lipstick ain't enough: beyond color matching for in-the-wild makeup transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13305–13314 (2021)
69. Kips, R., et al.: Weakly supervised color aware GAN for controllable makeup transfer. In: European Conference on Computer Vision, pp. 280–296. Springer (2020)
70. Zhu, J.-Y., et al.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)
71. Chang, H., et al.: Pairedcyclegan: asymmetric style transfer for applying and removing makeup. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 40–48 (2018)
72. Li, T., et al.: Beautygan: instance-level facial makeup transfer with deep generative adversarial network. In: Proceedings of the 26th ACM International Conference on Multimedia, pp. 645–653 (2018)
73. Zhang, H., et al.: Disentangled makeup transfer with generative adversarial network. arXiv preprint arXiv:1907.01144, (2019)
74. Yuan, Q.L., Zhang, H.L.: RAMT-GAN: realistic and accurate makeup transfer with generative adversarial network. Image Vis Comput. 120, 104400 (2022). <https://doi.org/10.1016/j.imavis.2022.104400>
75. Han, D., et al.: Spatially-invariant style-codes controlled makeup transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6549–6557 (2021)
76. Chen, H.J., et al.: Beautyglow: on-demand makeup transfer framework with reversible generative network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10042–10050 (2019)
77. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, (2014)
78. Zhang, K., et al.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process. Lett. 23(10), 1499–1503 (2016). <https://doi.org/10.1109/lsp.2016.2603342>

**How to cite this article:** He, F., et al.: Makeup transfer: a review. IET Comput. Vis. 1–14 (2022). <https://doi.org/10.1049/cvi2.12142>