J. Vis. Commun. Image R. 85 (2022) 103464

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Full length article

Facial makeup transfer with GAN for different aging faces☆,☆☆

Sen Fang [a,1], Mingxing Duan [a,*,1,2], Kenli Li [a,*,1], Keqin Li [a,b,1,3]

[a] School of Information Science and Engineering, Hunan University, China
[b] Department of Computer Science, State University of New York, USA

ARTICLE INFO

Keywords:
Domain
Facial aging
GAN
Makeup

ABSTRACT

Facial aging is widely used in criminal tracking and the search for lost children. If the aging face is made up, it will greatly affect the discrimination of the tracking system. Therefore, the research on the makeup of different aging faces is extremely important. Existing studies have achieved a good transition from the non-makeup domain to the makeup domain in facial makeup transfer. But few studies involve the transfer of facial makeup at different ages. In addition, existing datasets rarely contain both age and makeup attributes, which make the transfer of facial makeup for different ages full of challenges. To solve the above problems, we propose a learning framework, called AM-Net, which can realize facial makeup transfer for different ages while protecting identity information. AM-Net is composed of two sub-network modules: Aging-Net and Makeup-Net. AM-Net first learns the aging mechanism of faces through Aging-Net, and then, it feeds the learned aging mode to Makeup-Net. After that, AM-Net trains Makeup-Net to realize the mapping relationship between the non-makeup domain to the makeup domain and transfer the makeup style to the face of the non-makeup. Throughout the network, multiple losses are applied to ensure AM-Net preserve information about the identity, background, etc. Extensive experiments are conducted on different datasets with different state-of-the-art methods, which prove the effectiveness of AM-Net.

1. Introduction

1.1. Motivation

Facial aging is widely used in daily life, such as searching for missing children. However, the results of facial aging are easily affected by the external environment, and facial makeup has an extremely important effect on the results of facial aging. If a person has makeup, he/she can easily be disguised as very young or aging, which leads to low accuracy in the recognition of the aging face. As shown in Figs. 1 and 2, for most cosmetics, people look younger by applying facial makeup. Especially when people are middle-aged or old, the effect of makeup on facial aging becomes more obvious. That means that facial makeup has a very subtle effect on facial aging and the aging face after makeup will cover up facial wrinkles and other aging features, which will lead to inaccurate facial aging effects. In addition, there are few studies on the effect of facial makeup on facial aging. Therefore, the research on the mutual influence between facial aging and facial makeup is full of challenges. There are also extremely important challenges in studying facial makeup for different aging faces.

Makeup has become one of the common ways people beautify their faces in our daily life. Makeup mainly achieves facial beautification by concealing facial blemishes. However, the effect of makeup on facial beautification varies from the style of makeup to the way you utilize it. When customers are in a clothing store, they always try to wear various clothes before selecting the one that satisfies them. Similarly,

☆ This work was supported in part by the National Key-Research and Development Program of China under Grant No. 2020YFB2104003, in part by the National Outstanding Youth Science Program of National Natural Science Foundation of China under Grant 61625202, in part by the International Cooperation and Exchange Key Program of National Natural Science Foundation of China under Grant 61860206011, in part by the Shenzhen Excellent Technological and Innovative Talent Training Foundation under Grant RCBS20200714114941176, in part by the Science and Education Joint Project of Natural Science Foundation of Hunan Province under Grant 2020JJ7056. This paper is funded by the Hong Kong Scholars Program under Grants XJ2020032.

☆☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding authors.
E-mail addresses: sen.fang@hnu.edu.cn (S. Fang), duanmingxing@hnu.edu.cn (M. Duan), lkl@hnu.edu.cn (K. Li), lik@newpaltz.edu (K. Li).

[1] Sen Fang, Mingxing Duan, Kenli Li, and Keqin Li are with College of Information Science and Engineering, Hunan University, Changsha 410082, China.

[2] Mingxing Duan is also with the Shenzhen Institute, Hunan University, Shenzhen, 518063, China; he is also with the Department of Computing, Hong Kong Polytechnic University, Hong Kong, China.

[3] Keqin Li is also with Department of Computer Science, State University of New York, New Paltz New York 12561, USA.

https://doi.org/10.1016/j.jvcir.2022.103464
Received 3 February 2021; Received in revised form 30 December 2021; Accepted 12 February 2022
Available online 10 March 2022

Fig. 1. Examples of facial makeup transfer for people of different ages.

Fig. 2. Examples of the effects between facial aging and makeup.

facial makeup transfer provides a convenient way to help users choose their preferred makeup style. Face makeup style transfer is to let the original image learn other face makeup styles while maintaining the key attributes of the original image.

With the rapid development of image-to-image translation [1–3], makeup styles vary from person to person and are required at the instance level. Traditional makeup transfer schemes [4–6] are mostly based on physical operations, and recently, some methods are used to transfer makeup style based on deep neural networks [7]. In terms of makeup style transfer, BeautyGAN [8] successfully applied pixel-level histogram loss in local areas by integrating global domain-level loss and realized instance-level facial makeup transfer, which achieved remarkable results. However, BeautyGAN just implements makeup for an age of face shown in the current image but cannot solve the makeup problem for different age groups of the same face. For example, an 18-year-old girl finds herself beautiful when she puts on makeup. She suddenly wonders if she still looks more beautiful when she is old. Therefore, it remains a challenging problem to achieve facial makeup for the same person at different ages.

In terms of face aging, people have conducted a lot of research [9–11] on face aging in the early days, but due to the long-term lack of training samples for specific groups of people [12–15], the problem of face aging has become a challenging task. In recent years, many studies have made great progress in facial aging, and the effects of facial aging have become more and more realistic. The fidelity of facial aging and the protection of personal identity [16–18] are the basic premise of facial aging. Early aging methods commonly used include physical model-based methods and prototype-based methods. Physical model-based methods [19–21] simulate facial changes by modeling biological facial features such as skin and wrinkles. These methods often require a large amount of data, and their processes are complicated and computationally expensive. The prototype-based method [22,23] divides the dataset according to age groups and implements facial aging by learning differences across age groups. This method will lose lots of facial information, leading to unrealistic facial aging and even severe ghosting artifacts and other phenomena.

With the emergence of methods based on Generative Adversarial Neural Networks (GANs) [24], more and more high-quality images are generated, showing excellent capabilities in image generation [24–27]. Conditional Generative Adversarial Networks (CGANs) [28] is an improved method of generating adversarial networks. It implements conditional generative models by adding additional information. The application of CGANs in facial aging is also very successful, but most of this method requires paired training data and does not consider subtle changes such as facial expressions. There is still much room for improvement in the performance of facial aging.

To solve the above problems, we propose AM-Net, a generative adversarial neural network for facial makeup transfer that can learn from different ages of the same person. Although there are many researches on face aging, few of them involve makeup-based aging. In addition to solve the problems of facial aging and makeup, AM-Net also pays attention to the changes and results of facial aging after makeup, and explores the effects of aging after makeup.

1.2. Contributions

This paper proposes a facial makeup transfer learning method AM-Net for different ages. The network is composed of two sub-networks: facial aging network Aging-Net and facial makeup transfer module Makeup-Net.

The Patch-AutoEncoder module in Aging-Net is to learn facial aging characteristics. In the objective function of Aging-net, we adopt the perceptual loss to protect the character's identity, and texture loss to make the generated face look more natural and real. To distinguish between the generated face and the real face in the target age, we use the Estimator to evaluate the generated face.

Makeup-Net follows the CycleGAN. First, we utilize a generator to learn the features of the makeup domain and non-makeup domain and then learn the transfer process from the non-makeup domain to the makeup domain through the discriminator. To protect information such as the character's identity, we adopt perceptual loss and cycle consistency loss. The adversarial loss is to distinguish the generated image from the real sample image. We also utilize the pixel-level histogram loss calculated for different facial regions (such as eyes, lips, and face) to achieve instance-level transfer.

AM-Net learns the facial aging mechanism on the cross-age face recognition and retrieval dataset (CACD) and the Morph-II dataset and then uses it on the makeup dataset for makeup transfer. Experimental results have proved that AM-Net can well achieve facial makeup for the same person at different ages.

The rest of this paper is organized as follows: Section 2 presents closely related works. Section 3 illustrates the proposed method in detailed. Subsequently, we analyze the experimental results in Section 4 and the paper is concluded in Section 5.

2. Related work

2.1. Generative adversarial networks

Generative Adversarial Neural Network (GANs) [24] consists of two parts: generator and discriminator. The generator learns the feature distribution of real dataset and generates the feature distribution as close to the real dataset as possible. The discriminator is to distinguish the data distribution generated by generator from the real data distribution. In this way, generator and discriminator are in a dynamic fighting process. When this state is in balance, generator can generate very realistic data distribution. The objective function of the original GANs is as follows:

$$\min_G \max_D \mathbb{E}_{x\sim p_{data}(x)}[\log(D(x))] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]. \quad (1)$$

Due to the inconsistent convergence speed of the generator and the discriminator during the training process of GAN, and its training is unstable, some improved GANs have been proposed. Radford

et al. [29] designed a deep convolutional GANs, which combined CNN and GAN well. This method makes improvements to the structure of the convolutional neural network, which improves the quality of samples and the speed of convergence. Recently, Salimans et al. [30] presented an improved GAN that allows the model to perform better when generating high-resolution images.

2.2. Face aging

In terms of facial aging, the classical aging methods are divided into two parts: physical-based models [19–21] and prototype-based methods [22,23]. The physical model-based method simulates the aging mechanism of facial features, such as skin and wrinkles, through parametric or non-parametric learning. However, to learn this aging mechanism, this method requires a face dataset with a longer age span, and the calculation process of these algorithms is quite complicated. Wu et al. [31] simulated wrinkles and skin aging in facial animation by imitating the elastic process of connective adipose tissue between skin and muscle. The prototype-based approach groups the dataset according to age groups and then uses the average face of each group to construct the process of facial aging. The facial aging generated by the average prototype used in this method may lose the character identity, and the aging effect is unreal, and even severe ghosting artifacts may appear. Wang et al. [32] developed a method that uses super-resolution to simulate the effects of adult facial aging. Kemelmacher et al. [10] found a prototype-based solution, called conditional GAN (CGAN) [28] which has enabled researchers to achieve better and better results in facial aging and other aspects. Liu et al. [33] constructed Contextual Generative Adversarial Nets, which applies a conditional conversion network to simulate the aging process. Zhang et al. [34] established a CAAE framework. The framework adopts the encoder to map the face to the latent space, and then utilizes the deconvolution generator to project the vector to the face manifold conditional on age. Finally, two discriminators are applied to force the generation of the aging face. The Identity-Preserved Conditional Generative Adversarial Networks (IPCGANs) network model proposed by Wang et al. [35] utilizes perceptual loss to preserve the identity information, and adopts age classifiers to generate faces in the correct age group to achieve a good face aging effect. Duan et al. [36] proposed a new type of multi-attribute tensor correlation neural network (MTCN) for face attribute prediction, and they established a hierarchical prediction system called tensor correlation fusion network (TCFN) [37] for attribute estimation. Shu et al. [38] used a set of age-group dictionary models to express specific and personalized aging processes. Later, they [39] decomposed a personalized age progression (BDL-PAP) method based on two-layer dictionary learning to improve the performance of synthetic aging faces. Sun et al. [40] proposed a Label Distribution Guided Generative Adversarial Network (ldGAN) to deal with facial aging. Liu et al. [41] established a novel context generation adversarial network (C-GAN), which simulates the aging process through a conditional transformation network and two discriminative networks. Shi et al. [42] proposed a novel Conditioned-Attention Normalization GAN (CAN-GAN) for age synthesis by leveraging the aging difference between two age groups to capture facial aging regions with different attention factors. Duan et al. [43] decomposed DEF-Net, which can learn different facial expressions across different datasets and generate corresponding aging images. In addition, they constructed AR-Net [44] to explore the aging/rejuvenation (AR) features of human faces for age estimation. And they also established a black box attack method [45] called MBbA. This method encodes the input image and its target category into the associated space, and each decoder finds a suitable attack area from the image through the designed loss function, and then generates effective adversarial samples.

2.3. Style transfer

In terms of image-to-image translation, related research [1–3,46,47] has also achieved remarkable results. The Pix2Pix model introduced by Zhu et al. [1] solves the problem of image translation with pairs of data. The DiscoGAN discovered by Kim et al. [3] solves the cross-domain problem in the case of unpaired data. The Dual-GAN mechanism devised by Yi et al. [2] can be trained from two sets of unlabeled images from two domains. CycleGAN employed by Isola et al. [47] solves the problem of image translation under unpaired data. The StarGAN adopted by Choi et al. [46] utilizes only a single network model to effectively train the mapping between multiple domains from images in all domains.

3. Proposed method

As shown in Fig. 3, AM-Net is composed of two sub-networks: Aging-Net for learning the mechanism of facial aging and Makeup-Net for facial makeup transfer.

3.1. Aging-net

Aging-Net is mainly composed of PatchAutoEncoder and Estimator. PatchAutoEncoder consists of two parts: Encoder and PatchDecoder. The Encoder encodes the input image into the latent space z. The output z of $E(x) = z$ retains the advanced features of the input image x. PatchDecoder is responsible for extracting high-level features from z to generate specific faces. The Estimator discriminates the generated image by connecting and evaluating the classification obtained after convolution of the feature maps extracted from different layers.

3.1.1. PatchAutoEncoder

The emergence of GAN makes the generated images more realistic. The Encoder encodes the advanced features of the input image into a latent space. Then four residual blocks are used to preserve the spatial structure of the gradient. After that, the result is transmitted to PatchDecoder to realize the age transformation of the target spaces and we obtain the aging face image. PatchDecoder is used as a generator to synthesize the target image.

Generally, each layer in the network defines a nonlinear filter bank, and its complexity increases with the position of the layer in the network. Ref. [48] shows that cross-layer texture representation increasingly can capture the statistical attributes of images, so that the synthesized objects can retain more explicit information. While PatchDecoder synthesizes the target image, in order to generate the texture from a given source image, we first extract different features from each layer from the image in the PatchDecoder stage. Then, we compute a spatial summary statistic on the feature responses to obtain a stationary description of the source image.

When the model is vectorized, the layer with N_l different filters has N_l feature maps of each size M_l. These feature maps can be stored in a matrix $F_l \in \mathbb{R}^{N_l \times M_l}$, where F^l_{jk} is the activation of the j_{th} filter at position k in layer l.

To obtain detailed features such as the content and style of the original image, we design a PatchDecoder in the PatchAutoEncoder network to capture the character's texture and other information. For a given original image $\widetilde{x}$, we first calculate the activation of each layer l of the image $\widetilde{x}$ in the convolutional neural network. This feature acts on each layer of image generation which is composed of the correlation between each layer, and its feature correlation function is as follows:

$$G^l_{i,j} = \sum_k F^l_{ik} F^l_{jk}, \tag{2}$$

where G^l_{ij} represents the Gram matrix, $G^l \in \mathbb{R}^{N_l \times N_l}$, l represents each layer in the network, and G^l is the description of the features of each layer in the given texture network. This description corresponds to texture features in our network.

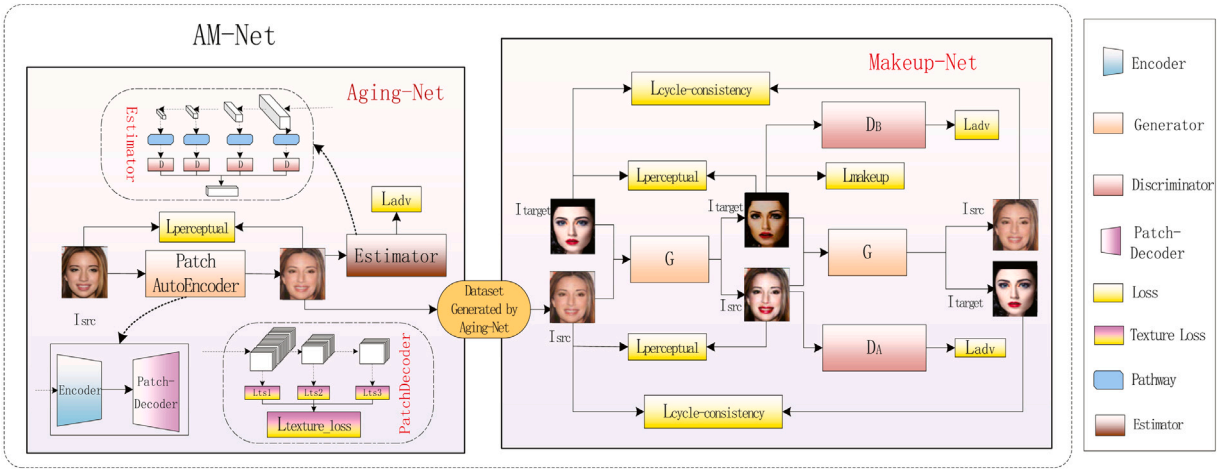

Fig. 3. The proposed AM-Net framework. AM-Net includes two networks: Aging-Net and Makeup-Net. Aging-Net is used to learn the aging characteristics of faces, and Makeup-Net is used to realize facial makeup transfer. In Aging-Net, PatchAutoEncoder is used to learn the facial features and aging mechanism of the original dataset, $L_{perceptual}$ is used to maintain the identity of the characters. In addition, $L_{texture\_loss}$ is used to ensure that the generated images are more textured and L_{adv} is used to distinguish the real image from the generated image. The Estimator evaluates the generated image. Makeup-Net adopts perceptual loss to maintain identity information, and utilizes cycle consistency loss, adversarial loss, and makeup loss to ensure that the network can generate makeup images.

We preserve the facial texture features by minimizing the mean square error between the original image's Gram matrix and the generated image's Gram matrix in each layer. The feature correlations of the original image x and the generated image $\widetilde{x}$, up to a constant scale, are given by the Gram matrix G^l, $\widetilde{G}^l$, respectively. The loss contributed by the l layer is:

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - \widetilde{G}_{ij}^l)^2. \tag{3}$$

After that, we can get the total texture loss. The objective function of the texture loss is as follows:

$$L_{texture}(x, \widetilde{x}) = \sum_{l=0}^{L} w_l E_l, \tag{4}$$

where w_l is a hyperparameter, which is the weight value contributed by layer l to the total loss. The standard error backpropagation can calculate the gradient of $L_{texture}(x, \widetilde{x})$ and E_l, where the gradient calculation formula of E_l is:

$$\frac{\partial E_l}{\partial \widetilde{F}_{i,j}^l} = \begin{cases} \frac{1}{N_l^2 M_l^2}((\widetilde{F}^l)^T(G^l - \widetilde{G}^l)), & \text{if } \widetilde{F}_{i,j}^l > 0; \\ 0, & \text{if } \widetilde{F}_{i,j}^l < 0. \end{cases} \tag{5}$$

In addition to capture the texture features of faces, PatchDecoder is also used to generate the faces. We adopt the transposed convolutional layer to restore the high-level features extracted from the Encoder into an image of the target age domain. To achieve this goal, we utilize the mean square loss function to replace the log-likelihood to make the generated data distribution as close to the real data distribution as possible. The objective of the generated image is:

$$L_{adv}(G) = \mathbb{E}_{x \sim p_{young}(x)}[(D(G(x)) - 1)^2]. \tag{6}$$

As a discriminator, D distinguishes the real image from the generated image, and G generates the target image. G and D train alternately until the dynamic balance is reached.

3.1.2. Estimator

The Estimator is used to discriminate the images, and these images include the original image, the aging image of the target, and the generated image. The main body of the Estimator adopts a pyramid structure with multiple discriminators and multiple levels of discrimination. To achieve this structure, we extract the feature maps of the second layer, the fourth layer, the seventh layer, and the tenth layer from the VGG-16, and then distinguish each layer to obtain the classification results.

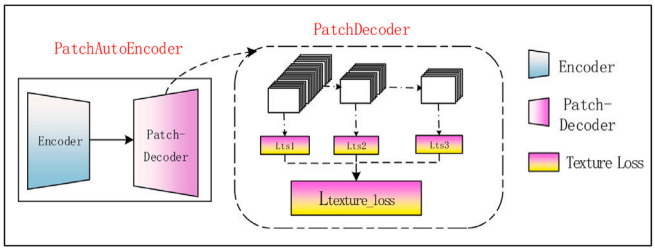

Fig. 4. The network structure of PatchAutoEncoder.

Finally, the classification results of each layer are spliced and compared with the real tags to get the final results of image discrimination by the Estimator. We denote the image distribution sampled from the original data as $x \in p_{young}(x)$ and the generated target image as $\widetilde{x} = G(x) \in \widetilde{p}_{old}(x)$. To distinguish the real image from the generated image, that is to make the generated data distribution as similar as possible to the original data distribution, namely $\widetilde{P}_{old} \approx P_{old}$, we define the objective function of the discriminator as follows:

$$\begin{aligned} L_{adv}(D) = &\frac{1}{2}\mathbb{E}_{x \sim p_{old}(x)}[(D(x) - 1)^2] + \\ &\frac{1}{2}\mathbb{E}_{x \sim p_{young}(x)}[(D(G(x)) - 1)^2 + (D(x) - 1)^2]. \end{aligned} \tag{7}$$

3.1.3. Identity protection module

How to preserve the personal identity information during the image generation process is very important. However, the adversarial neural network can only generate samples that conform to the real data distribution as much as possible, but the adversarial loss cannot preserve the identity information. To enable the generated face to maintain the identity information of the origin image, we introduce a perceptual loss in the objective function of Aging-Net:

$$L_{identity} = \sum_{x \in p_{src}(x)} (\|d(x) - d(G(x))\|)^2, \tag{8}$$

where $d(\cdot)$ is the Euclidean distance between feature representations. The generated faces are made more realistic by measuring the mean square error between the features of the original image and the features of the generated images. Therefore, images generated by the generator do not lose the identity information.

3.1.4. Optimization (Aging-Net)

As a sub-network in AM-Net, Aging-Net learns the aging mechanism of faces and transfers the learned aging mechanism to Makeup-Net. In Aging-Net, the optimization goals of the entire system are:

$$L_D = L_{adv}(D), \tag{9}$$

$$L_G = \lambda_{adv} L_{adv}(G) + \lambda_{id} L_{identity} + \lambda_{tex} L_{texture}, \tag{10}$$

where λ_{adv}, λ_{id}, λ_{tex} are hyperparameters to balance the relative importance of each module. λ_{id} is used to regulate the degree of identity information protection and λ_{tex} is applied to adjust the texture condition during face generation.

3.2. Makeup-net

The Makeup-net is to realize makeup transfer of faces, aiming to transfer a given makeup style to the face without makeup. For a given dataset A without makeup and a makeup dataset B, we sample the original image $I_{src}(x) \in A$ and the target image and $I_{tgt}(t) \in B$ from these two datasets. Then through the network model learning $(I^{tgt}_{src}(x), I^{src}_{tgt}(t)) = G(I_{src}(x), I_{tgt}(t))$, the makeup image $I^{tgt}_{src}(x)$ is obtained, where $I^{tgt}_{src}(x)$ represents the makeup style of the target data $I_{tgt}(t)$ is transferred to the original data $I_{src}(x)$. In addition, the identity features of the original image are preserved. The generator G is used to generate the target image, and the discriminator distinguishes the generated image from the original image.

3.2.1. Adversarial loss

The generator learns the data distribution $p_{data}(x)$ of the training sample to simulate a data distribution which is close to $p_{data}(x)$. The discriminator is to distinguish between the real data and the data generated by the generator. When the generator and the discriminator are in a mutually restrictive dynamic balance, the final data distribution $\tilde{x}$ generated by the generator will be close to the real data distribution x. For the adversarial loss between the data $I_{src}(x)$ sampled from the dataset A that has not been made up and the target image $I^{tgt}_{src}(x)$ generated by the generator, the objective function is as follows:

$$\begin{aligned} L_{adv}(D_A) = {} & \mathbb{E}_{I_{src}(x)}[(D_A(I_{src}(x)))^2] + \\ & \mathbb{E}_{I_{src}(x), I_{tgt}(t)}[(D_A(G(I_{src}(x), I_{tgt}(t))) - 1)^2]. \end{aligned} \tag{11}$$

Similarly, to distinguish the image generated by D_B from the dataset sampled from the makeup dataset B, the objective loss function is calculated as follows:

$$\begin{aligned} L_{adv}(D_B) = {} & \mathbb{E}_{I_{tgt}(t)}[(D_B(I_{tgt}(t)))^2] + \\ & \mathbb{E}_{I_{src}(x), I_{tgt}(t)}[(D_B(G(I_{src}(x), I_{tgt}(t))) - 1)^2]. \end{aligned} \tag{12}$$

We adopt the adversarial loss to ensure that the network can generate realistic dataset.

3.2.2. Identity protection

The Makeup-Net we proposed adopts a dual input/output architecture. A single generator G learns the mapping relationship between makeup domains and non-makeup domains, which realizes the transfer of makeup styles to images. To make the generated images more realistic by preserving information such as identity features and background, we impose perceptual loss and periodic consistency loss to constrain them.

We use the VGG network to extract high-level features of the image. For the input image x, we use $F_l \in \mathbb{R}^{(H_l \times W_l \times C_l)}$ to denote the l_{th} layer of feature information extracted from the VGG network. H_l, W_l, and C_l respectively represent feature mapping the height, width and the number of feature maps. The objective function of the perceptual loss is written as follows:

$$L_{perceptual} = \frac{1}{H_l \times W_l \times C_l} \sum_l E_l, \tag{13}$$

$$\begin{aligned} E_l = \sum_{i,j} (& [F^l_{i,j}(I_{src}(x)) - F^l_{i,j}(I^{tgt}_{src}(x))]^2 + \\ & [F^l_{i,j}(I_{tgt}(t)) - F^l_{i,j}(I^{src}_{tgt}(t))]^2), \end{aligned} \tag{14}$$

where $F^l_{i,j}$ represents the activation of the filter at position (i, j) in the l_{th} layer. To make the generated image as close to the original image as possible, we adopt cycle consistency loss to make the generated image retain the background, content and other information of the images in addition to the features after conversion. We utilize the $L2$ norm to measure the distance between the generated image and the original image. The loss functions are:

$$L^{recon}_{src} = \mathbb{E}_{I_{src}(x), I_{tgt}(t)}[\| I_{src}(x) - \widetilde{I}_{src}(x) \|_2], \tag{15}$$

$$L^{recon}_{tgt} = \mathbb{E}_{I_{tgt}(t), I_{tgt}(t)}[\| I_{tgt}(t) - \widetilde{I}_{tgt}(t) \|_2], \tag{16}$$

where $(\widetilde{I}_{src}(x), \widetilde{I}_{tgt}(t)) = G(G(I_{src}(x), I_{tgt}(t)))$, $(\widetilde{I}_{src}(x), \widetilde{I}_{tgt}(t))$ is the restoration of the target image by the generator, and its goal is to make $(\widetilde{I}_{src}(x), \widetilde{I}_{tgt}(t)) \approx (I_{src}(x), I_{tgt}(t))$. In the entire cycle consistency loss, we use the $L2$ norm to accelerate the model convergence. Then, the total cycle consistency loss is:

$$L_{cycle} = L^{recon}_{src} + L^{recon}_{tgt}. \tag{17}$$

3.2.3. Instance-level makeup transfer

To achieve instance-level makeup transfer, we utilize a histogram matching approach that introduces histogram loss at the pixel level. Since we tend to apply makeup at specified locations such as cheeks, lips, etc., when applying makeup, we extract only three regions of the image: face region, lip region, and eye region, and then calculate the local histogram loss of each part separately and finally get the total makeup loss. We apply the local histogram loss to the three areas of the face, lips, and eyes. The loss function is:

$$L_i = \| I^{tgt}_{src} - HM(I^{tgt}_{src} \cdot FP_i(I^{tgt}_{src}), I_{tgt} \cdot FP_i(I_{tgt})) \|_2, \tag{18}$$

where $i \in \{face, lips, eyes\}$. After that, the makeup loss is:

$$L_{makeup} = \lambda_f L_{face} + \lambda_l L_{lips} + \lambda_e L_{eyes}, \tag{19}$$

where λ_f, λ_l, λ_e are the weight coefficients of each calculation area.

3.2.4. Optimization (Makeup-Net)

Makeup-Net is used to realize the transfer process of facial makeup. AM-Net first receives the aging mechanism learned from Aging-Net, then studies the mapping relationship between the non-makeup domain to the makeup domain in Makeup-Net, and transfers the makeup style to the non-makeup face. The entire AM-Net realizes the facial makeup transfer for different ages. The optimization of the objective function of Makeup-Net is:

$$L_{adv}(D) = L_{adv}(D_A) + L_{adv}(D_B), \tag{20}$$

$$\begin{aligned} L_{adv}(G) = {} & \lambda_{adv} L_{adv} + \lambda_{per} L_{perceptual} + \\ & \lambda_{cycle} L_{cycle} + \lambda_m L_{makeup}, \end{aligned} \tag{21}$$

where λ_{adv}, λ_{per}, λ_{cycle}, λ_m are hyperparameters that measure the relative importance of each item.

4. Experimental results

4.1. Dataset

After evaluating the purpose of our experiment and the significance of the research, we use the cross-age face recognition and retrieval dataset (CACD) [50] as the facial aging dataset used by AM-Net to train Aging-Net. The dataset (CACD) contains 163,446 images of 2,000 celebrities, ranging in age from 14 to 62. From the results of some experiments, we find that if the face is aging at a specific age, the

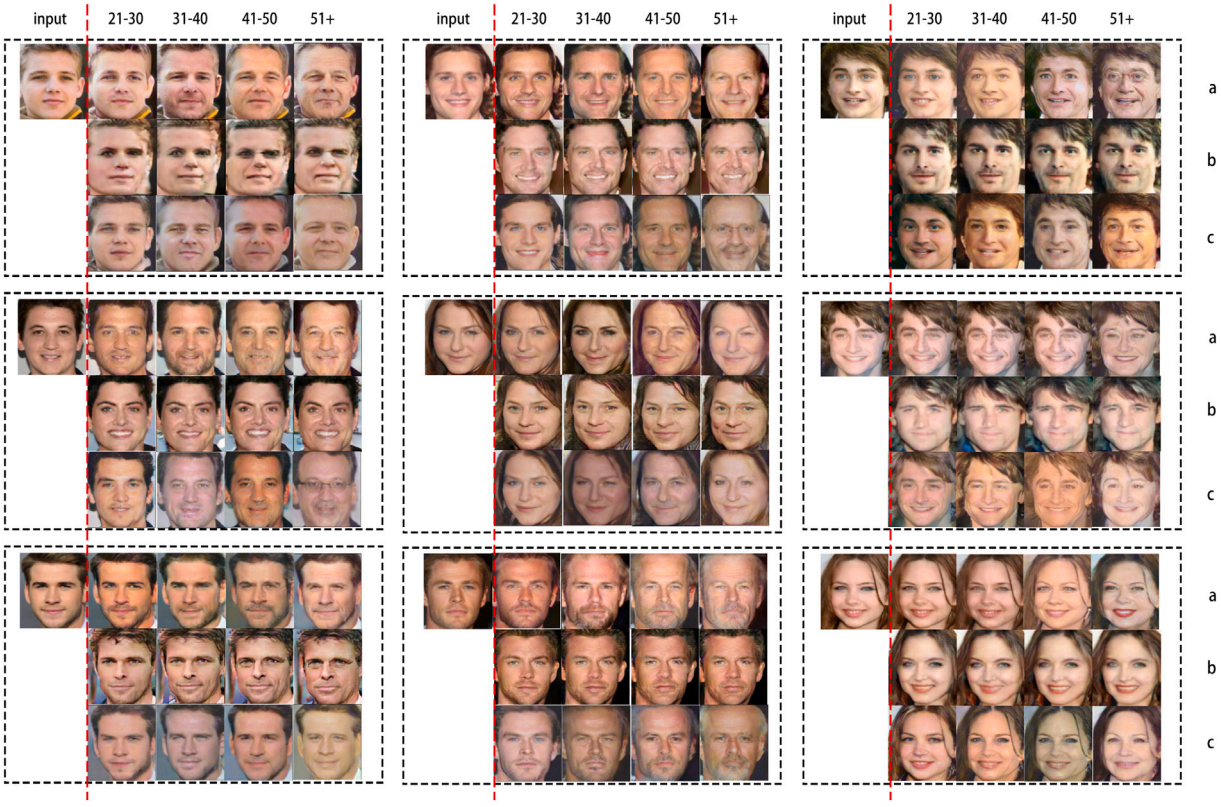

Fig. 5. Aging face generated by different methods. Each black dashed box represents the same person, the left side of the red dashed line is the test image, whose age range is about 16–18 years old, and each column on the right side corresponds to the aging effect of different age groups, and the age range is 20–62 years old. Ten years old is a division of age group. The lines a, b, and c in each black box correspond to the faces generated by Aging-Net, CAAE [34], and a pyramid architecture of GANs [49].

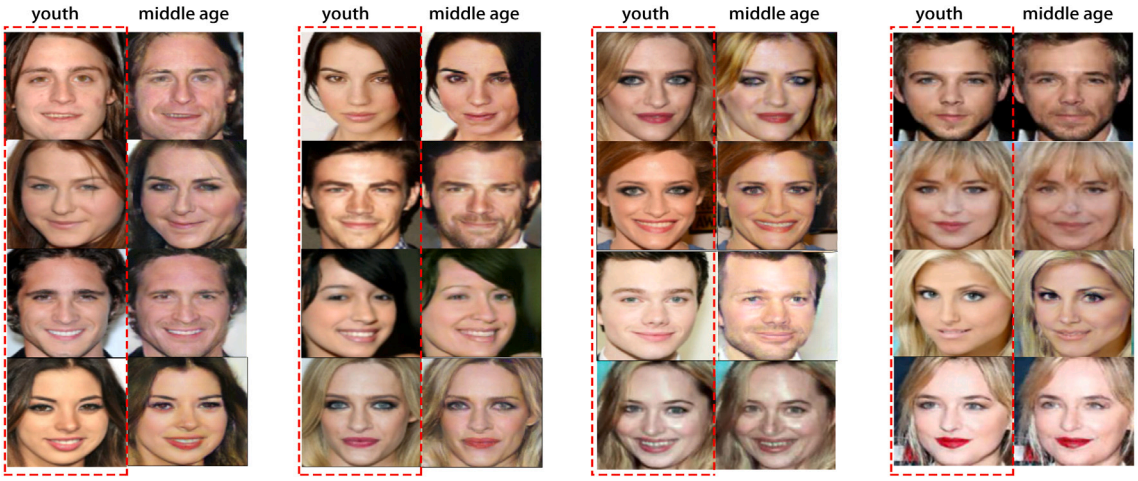

Fig. 6. Examples of facial aging. The red dotted box represents the young group to be tested, and the next to the right is the result of the target age group (middle-aged).

training process is difficult to converge, mainly because the facial features with small age differences are too similar to be distinguished during training. We classify celebrities in the 14–18 years old into the young group as the input to the original dataset; take 20–30, 31–40, 41–50, and 51+ years old as the target age group respectively. We randomly select 90% of the images for training and the remaining 10% for testing. We use the Makeup dataset published by Li et al. [8] to train the facial makeup transfer module of the network. We crop all the images of the dataset in the center of the image and then resize images to 224 × 224. Finally, we use the test dataset portion of the CACD dataset and the Morph-II dataset for performance testing of the model.

4.2. Implementation of AM-Net

Network Architecture. We design PatchAutoEncoder to capture the aging mechanism of faces. The Encoder maps the input image to the latent space, which is composed of three convolutional layers. Then we use four residual blocks [51] to preserve the spatial structure of the gradient as much as possible. Then through the PatchDecoder to achieve the change of the target age. PatchDecoder consists of three transposed convolutional layers. All convolutional layers are followed by Instance Normalization and ReLU nonlinear activation functions. Estimator consists of four discriminators, which concatenate the results of the four discriminators classified separately. Each discriminator is

consists of three parts: convolutional layer, batch Normalization, and LeakyReLU, which respectively distinguish the extracted feature maps of the second layer, the fourth layer, the seventh layer, and the tenth layer. The generator G adopts a dual-input–output structure. First, the two input ends are concat, down-sampled, and transferred to several remaining blocks. After that, two target images are output through two independent up-sampling processes. For the two discriminators D, we adopt the discriminative output structure of PatchGANs to fuse the local image features with the overall image features.

Training Details. We flip the image horizontally and vertically to enhance the dataset and set the experimental batch size to 16. We use Adam with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ to optimize all models and set the learning rate to 0.0001. The decay learning rate is updated every 1000 steps, and all models are trained for 50,000 epochs. The parameters of the aging mechanism learning stage are set as: $\lambda_{adv} = 50$, $\lambda_{id} = 0.5$, $\lambda_{tex} = 50$. In the facial makeup transfer phase, we set the training batch to 1, and set the other parameters to $\lambda_{adv} = 5$, $\lambda_{per} = 0.05$, $\lambda_{cycle} = 10$, $\lambda_m = 1$. λ_f, λ_l, and λ_e are all set to 2. We use pre-trained VGG-16 to make the identity of the generated faces be not lost. We normalize the pixel values of the image and limit the numerical normalization range to [0, 1].

4.3. Qualitative comparison

4.3.1. Results of Aging-Net

We first extract people under 20 years old from the dataset as the young group, and then select people around 45 years old as the middle-aged group, with an age span of about five years within each of the two groups. Fig. 6 shows the aging images generated by Aging-Net. The aging faces in the figure are the test dataset for the young group, and the images for the middle-age group are the target image generated by Aging-Net. As can be seen from Fig. 6, Aging-Net can learn the aging mechanism of the face well, and its effect is mainly on the texture characteristics of facial wrinkles and skin texture. Moreover, our experimental results present a visually almost realistic aging effect. After that, we train the CAAE code released by Zhang et al. [34] and the code released by Yang et al. [49]. From Fig. 5, we can learn that the CAAE model has a high probability of losing the identity information of the character and making the generated aging face too fake. Besides, the aging effect generated by this model only concentrates on the wrinkles at the corners of the mouth, which leads to the generated aging pattern is too single. The aging images generated by Yang et al.'s [49] model have some unobvious aging. The model produces a face that looks more like an aging film laid over the face, which looks unrealistic and carries the risk of losing identity. Aging-net can well maintain the identity information, the aging effect is more visible, natural, and has fewer artifacts. In addition, the generated aging image has higher quality.

The effectiveness of texture loss. In Aging-Net, we design PatchAutoEncoder to capture information such as the facial texture to make the generated image more natural and lifelike. To speed up the model training, we use the VGG-16 network structure to replace the original PathchDecoder calculation method of texture loss. As shown in Fig. 7, the texture generation method used in column b is more realistic than the unused effect in column a, and the image without texture loss will have visible character distortion and artifacts. In column a, the images generated by the model without texture loss are not too visible in the aging effect, and the generated faces are also blurry. Because column b uses texture loss to capture the detailed features such as the texture of the face, the generated image shows a good effect in terms of the aging effects, characters' identity protection, or the degree of facial realism.

Robustness verification. The aging effect of the face is often affected by external factors. For example, when the light is intense, the face can appear particularly harsh, which can cause an inability to recognize the face when aging. Similarly, in a dark environment, aging a dark face is also challenging. Also, there will be some obstructions

Fig. 7. Examples for verifying the effectiveness of texture loss. The red box on the left is the input image, and the two columns on the right are the aging effects of different methods. Column a is the result without texture loss, and column b is the result with texture loss to make the generated face more natural.

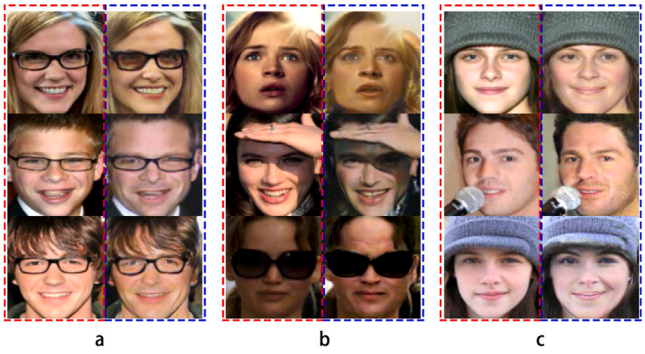

Fig. 8. Examples of facial aging in different situations. Column a is the aging of the face when wearing glasses; column b is the aging of a person under intense light and dark environment; column c is the aging of the face under other obstructions such as hats and microphones.

on the face, such as wearing glasses, wearing a hat, and a small area of the face blocked by other objects. These situations can also have a significant impact on facial recognition as well as aging works. We randomly select and test some images to verify the stability and robustness of the AM-Net framework. Fig. 8 shows the effect of facial aging in different situations. Column a is the case with glasses, and Column b is the aging under intense light and low light. The upper and lower rows of column c have a hat, and the middle row of column c has a microphone beside it. It is shown in Fig. 8 that Aging-Net can learn the facial aging mechanism of a character without being influenced by the external environment and other factors and can make realistic aging effects.

Comparison with ground truth. To verify that the proposed framework can protect personality in terms of aging, we qualitatively compare the generated facial aging images with the real ground conditions. As can be seen from Fig. 9, the aging face generated by our proposed framework bears a striking resemblance to the real faces. These similarities are reflected in identity consistency and facial wrinkles, textures, other identities, and facial aging information.

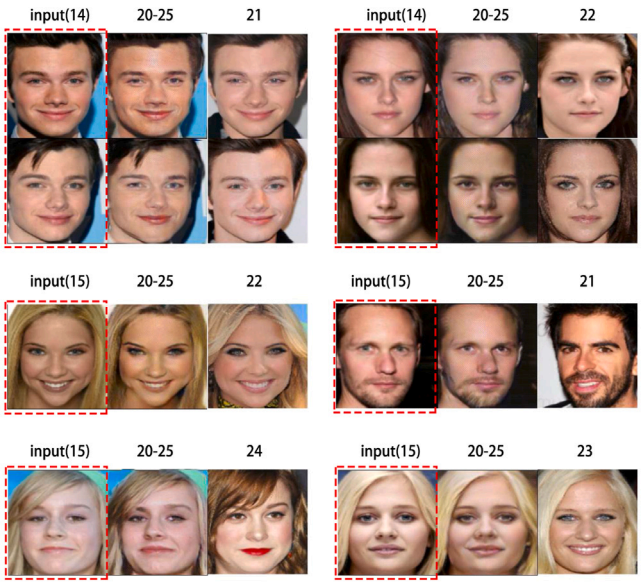

Fig. 9. Comparison with ground truth.

4.3.2. Results of Makeup-Net

Fig. 11 shows the effect of different styles of makeup when people are young and middle-aged. As can be seen from Fig. 11 that people look different in their makeup when they are young and middle-aged. The main reason is that with the increase of age, the facial muscles, skin, and other characteristics of faces will change slightly, and the larger the period, the more visual this change will be. AM-Net's goal is to achieve facial makeup transfer of different ages. As can be seen from Fig. 10 that some makeup can still hide facial aging features such as wrinkles on the corners of the mouth, which makes some people look younger after makeup. Also, some people appear more mature and stable after putting on makeup at different ages. In addition to the reasons for the increase in age, makeup also plays a large role. In conclusion, the effect of facial makeup varies not only from person to person but also from makeup style to makeup style. Fig. 16 shows the experimental results of AM-Net on the original uncropped image. It can be seen from the figure that AM-Net can perform facial aging and makeup well in large-size, high-pixel images. In comparison, AM-Net has a more obvious effect of facial aging on the extracted facial regions.

4.3.3. Makeup effects for different ages on different models

As shown in Fig. 13, we show the results of qualitative comparison with other models. We observe that the images produced by Guo et al. [52] have obvious artifacts and look unreal. The model of Liu et al. [53] can perform the transfer of facial makeup very well, but there will be obvious mismatches in the wheel library area of the face. Compared with the above models, AM-Net has considerable effects on the quality of the generated images and the transfer of makeup styles. Fig. 14 shows the aging effect of Aging-net on the original high-pixel image. However, we find some negative effects generated by Aging-Net. As shown in Fig. 15, when performing facial aging on a person's face, very few people have gender inconsistencies. After analysis and verification, we find that the main reason is that the faces of some people resemble the opposite sex, and the uneven ratio of male and female faces in the dataset may cause the features learned by the network to be more biased towards facial features with a high proportion of gender.

Qualitative evaluation. In Fig. 12, we show the comparison between AM-Net and other methods in facial makeup transfer. As can be seen that CycleGAN [47] can generate relatively realistic images, but the generated makeup effects are subtle and not particularly obvious. For example, there is only a shallow trace of makeup on the cheeks and lips of the face. But our AM-Net can generate high-quality images with the most accurate makeup style. The images generated by AM-Net not only have better effects on eye shadows, cheeks, and lips but also retain other non-makeup-related elements such as hair, clothes, and backgrounds.

4.4. Quantitative comparison

4.4.1. User study evaluation

To make the experimental results more objective and fair, many researchers adopt the user evaluation method as a quantitative assessment of the experiment. To quantitatively evaluate AM-Net, we randomly interviewed 60 volunteers to evaluate the target images we generated. Their tasks include three parts: face authentication, authenticity evaluation of facial aging, and evaluation of makeup effects. We randomly choose 100 test images and 30 reference makeup test images between 16–20 years old from the dataset. Aging-Net generates each image separately for four different age groups. Then we use these images to generate images of four different age groups by using the CAAE network model [34] and the model of Yang et al. [49]. These generated images are used for volunteers to evaluate the effects of face aging experiments. We randomly select 400 images from these 1,000 test images, and 30 reference makeup test images from the dataset. After the AM-Net test, a total of 32,000 facial makeup images of different ages are obtained from the test images and then used with the makeup images generated by CycleGAN [47] to allow volunteers to evaluate the effect of facial makeup.

Facial Aging Evaluation. We ask volunteers to give the age range corresponding to each aged face. Then we calculate the proportion of the results that the volunteers given the same age range as our calibrated age range.

Makeup Evaluation. We show the volunteers four images each time, namely the original image after facial aging, the reference makeup image, and two randomly generated images after makeup transfer. These images are generated by AM-Net as well as CycleGAN [47], respectively. We ask the volunteers to give the degree of recognition of the makeup effect of each image (percentage system). We average the total recognition obtained for each method separately, and the value is as the level of recognition of each method by the volunteers.

Face Verification. We ask volunteers to evaluate whether the generated face can preserve identity information and whether it belongs to the same person. For each input image, we generate four aged images given different age labels. We divide into five pairs: (input, age1), (age1, age2), (age2, age3), (age3, age4), (a randomly selected generated image, a randomly selected image of other persons). The first four pairs are to verify whether the generated image is the same as the input. The last pair is to verify whether the generated image is the same as the other person. Then we ask the users to do the face verification task and report the accuracy of the different methods. If the trained model fails to protect the identity information or generates the same aged face with different inputs, the facial verification score is low.

Table 1 shows the performance comparison of various methods. As can be seen from the table that our AM-Net has a higher performance than the baseline model CAAE [34] and Yang et al. [49] on the task of facial aging; and it performs better than CycleGAN [47] on the task of facial makeup transfer. In terms of facial verification, AM-Net has also achieved high performance, which shows that our method can generate high-quality facial makeup faces for different ages while protecting the identity information.

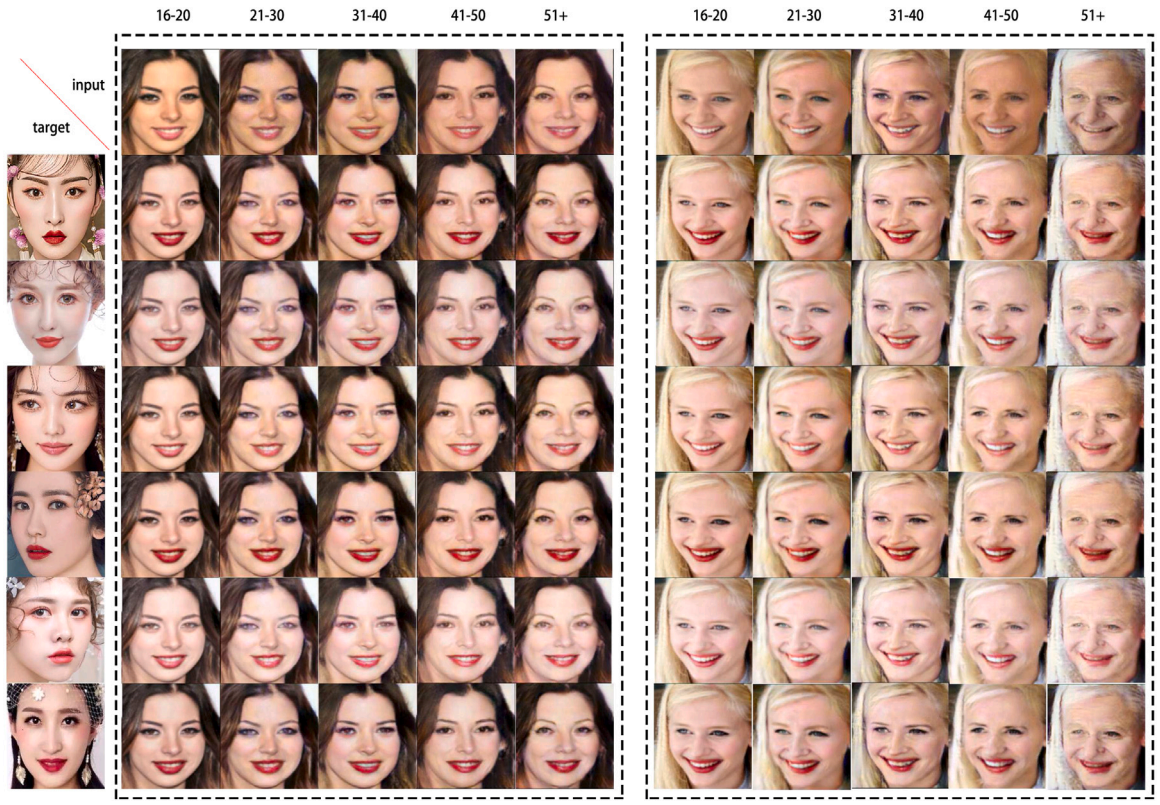

Fig. 10. The result of facial makeup transfer at different ages. The leftmost column is the target makeup picture; the result in the same black box is the transfer of facial makeup of the same person at different ages. In each frame, from left to right is the direction of facial aging, and from top to bottom represents different makeup styles.

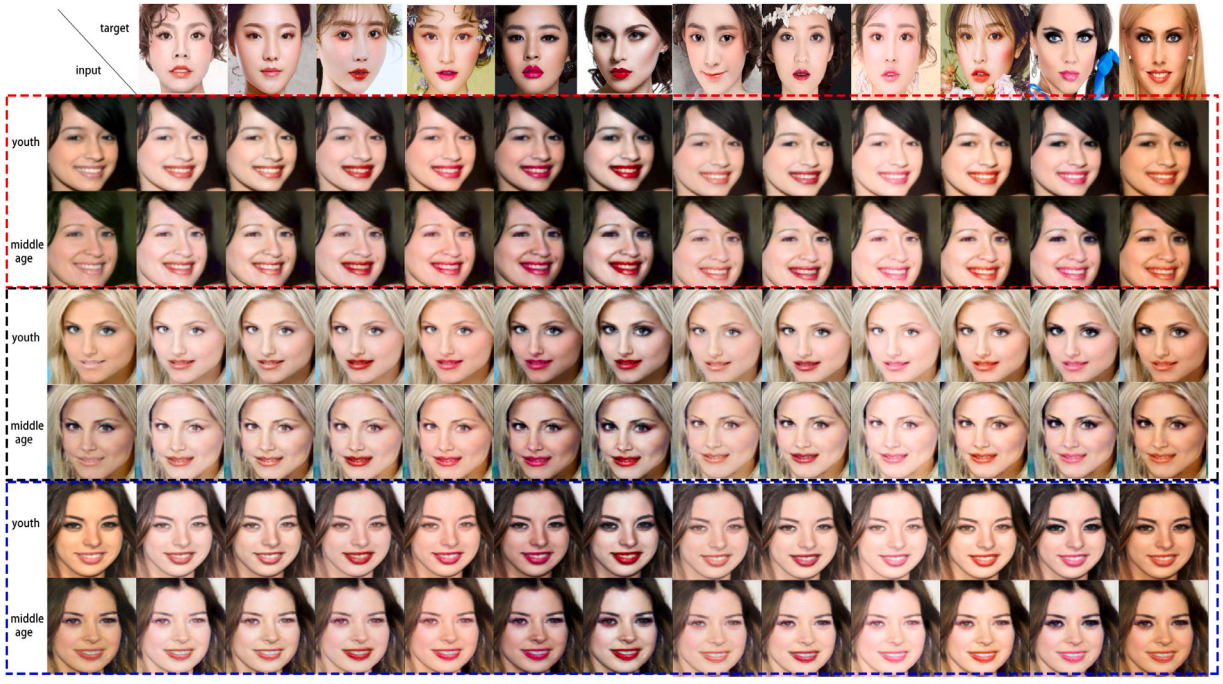

Fig. 11. Examples of AM-Net. The images in different color boxes represent different identity information. Each color box represents the result of different facial makeup transfers of the same person who is young and middle-aged.

4.4.2. *Face aging accuracy with synthesized dataset*

With the aging of the face, the corresponding age of the generated face is also increasing. Evaluating the generated images by users is just a universal evaluation method. To measure the age corresponding to the generated face more accurately, we adopt the online Analyze API of Face++ [55]. Passing the detected faces into the Analyze API, we can obtain the key points, age, gender, head posture, smile detection, glasses detection, face quality, and other information. But we only extract the age field returned by the API. We apply this tool to each synthesized face. We randomly select 3000 images from the

Fig. 12. Qualitative comparison. The red box is the input test image, the blue box is the reference makeup image, and the third and fourth columns are the makeup transfer results generated by CycleGAN and AM-Net.

Fig. 13. The makeup effects of different makeup models at different ages. The three rows of a,b,c below the red box are the makeup effects of AM-Net, Guo [52], and Liu [53] models respectively.

Fig. 14. The aging effect of Aging-Net on the uncropped, high-pixel original image.

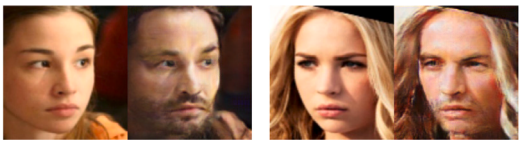

Fig. 15. The aging effect of Aging-Net on the uncropped, high-pixel original image.

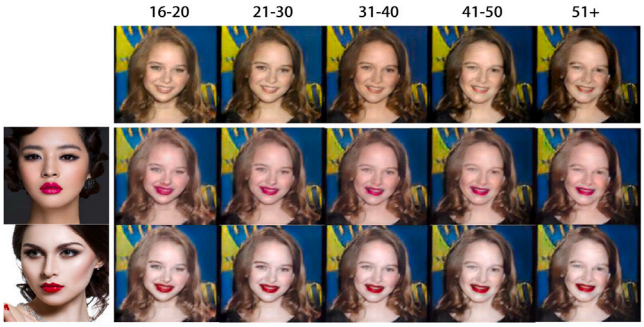

Fig. 16. The experimental effect of AM-Net on the original image.

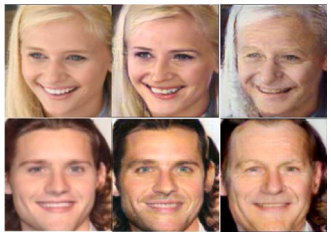

Fig. 17. The aging effect of AM-Net.

Table 1
Performance comparison with different methods.

	CAAE	Yang	CycleGAN	AM-Net
Facial aging	36.17%	52.16%	–	63.67%
Makeup effects	–	–	31.15%	68.85%
Face verification	56.24%	91.24%	–	94.13%

CACD dataset and Morph-II dataset each time and then generate four different age faces through AM-Net, a total of 12,000 images for age assessment. When calculating age assessment accuracy for each age group, we exclude the counts of faces that the API does not recognize, returned age fields with null values, and cases where the API request fails. After experiments are cross-validated five times, we calculate the approximate range of facial aging accuracy between the synthetic facial aging and the facial aging in the original dataset. We denote the age groups under 20 years old, 21–30 years old, 31–40 years old, 41–50 years old, 51 years old and above as Age Group 0, Age Group 1, Age Group 2, Age Group 3, Age Group 4. Our experimental results of facial aging accuracy include the assessment of facial aging generated by AM-Net and the evaluation of facial age in the original dataset. We can learn from Table 2 that the facial ages generated by AM-Net remain within the range of each age group, and the error range of each age group is within the corresponding range of the age range. Table 3 shows the ratio of the generated face and the real face that can be correctly matched by the detected age and the respective label. Because there is a big difference between the age of Face++ online evaluation and the age of the given label in the dataset, all the values are low while the overall trend can still see facial aging. Although there is a considerable deviation between the speed of facial aging and the actual age due to factors such as personal lifestyle and living environment,

Table 2
The average age range of each age group of generated faces.

Datasets	Age Group 1	Age Group 2	Age Group 3	Age Group 4
CACD	26.54 ± 3.23(6.16 ± 1.13)[a]	39.64 ± 2.31(9.51 ± 2.01)[a]	46.43 ± 5.73(9.26 ± 3.12)[a]	53.43 ± 6.37(11.48 ± 2.23)[a]
Morph-II	24.13 ± 5.18	37.23 ± 3.24	43.25 ± 6.13	61.23 ± 9.82

[a]The value in brackets represents the standard deviation.

Table 3
The proportion of the detected age and the respective label consistency.

		Age Group 0	Age Group 1	Age Group 2	Age Group 3	Age Group 4
CACD	Face Styles(%)	–	76.27 ± 9.23	43.57 ± 10.31	42.75 ± 6.73	53.43 ± 6.37
	Datasets(%)	31.07 ± 9.42	62.28+7.16	37.24 ± 9.87	39.56 ± 6.73	47.75 ± 8.34
Morph-II	Face Styles(%)	–	71.74 ± 6.87	69.35 ± 7.34	56.23 ± 9.71	85.45 ± 8.68
	Datasets(%)	22.68 ± 5.34	54.74 ± 6.73	53.23 ± 8.76	46.89 ± 5.67	72.59 ± 9.97

Table 4
The results of different evaluation schemes and different experimental methods.

	Test results on the SSR-Net model [54]				Volunteer evaluation			
	Face aging effects		Makeup effects		Face aging effects		Makeup effects	
	CACD	Morph-II	CACD	Morph-II	CACD	Morph-II	CACD	Morph-II
Method-1[a]	64.13%	71.41%	–	–	71.21%	69.98%	94.23%	96.36%
Method-2[b]	79.98%	81.32%	–	–	87.35%	89.21%	85.43%	88.82%

[a]Method-1: The network first learns the aging mechanism of the face, and then generates the results of facial makeup for different ages (Aging-Net + Makeup-Net).
[b]Method-2: The network first generates faces with different makeup styles, and then uses these makeup faces to generate face aging images (Makeup-Net + Aging-Net).

Table 5
The proportion of predicted facial age in each age group.

Face Aging											
		CACD					Morph				
		Under 20	21–30	31–40	41–50	51+	Under 20	21–30	31–40	41–50	51+
Method-1 (Aging-Net + Makeup-Net)	21–30	2.13%	93.12%	3.32%	1.21%	0.21%	2.61%	94.97%	2.12%	0.21%	0.08%
	31–40	0.37%	2.09%	88.60%	6.41%	2.53%	0.19%	3.97%	82.64%	9.46%	3.74%
	41–50	0.13%	0.73%	3.15%	86.60%	9.39%	0.01%	1.01%	1.21%	87.56%	10.21%
	51+	0.03%	1.68%	3.21%	6.13%	88.95%	0.20%	1.48%	3.95%	5.36%	89.01%

Table 6
The results of recognition on different datasets.

	CACD		Morph-II	
	OD[a]	GD[b]	OD[a]	GD[b]
Age Estimate	78.79%	75.34%	81.26%	79.16%

[a]Test dataset randomly selected from the CACD dataset and Morph-II dataset.
[b]Test dataset generated from AM-Net.

the trend is still relatively unchanged. As seen in Tables 2 and 3, AM-Net can learn facial aging patterns at other ages even though there is a difference between the evaluated age and the age corresponding to the label. The estimated age of the face synthesized by AM-Net is within the corresponding age range, and the proportion of accuracy is higher than the age of the real image. It is shown that the age of the AM-Net synthesized face fits well with the actual age and shows a steady increasing trend over time, which validates our method.

4.4.3. Accuracy of facial aging obtained by using predictive algorithms

We use a deep learning model to predict the age of the generated face. First, we randomly select 8000 images under the age of 20 from the CACD dataset and Morph-II dataset, and then generate four different age facial aging images through AM-Net. The original dataset generates a total of 40,000 images through AM-Net. We use the SSR-Net model [54] to verify the accuracy of facial aging. We define the accuracy rate as the ratio of the predicted age by SSR-Net to the age range of the face generated by AM-Net. We use 30,000 images for model training and the rest for testing. To ensure the fairness of the results as much as possible, we keep the size of the generated test dataset consistent with the original test dataset. Table 6 shows the results of SSR-Net's age prediction on the face in the original dataset and the dataset generated by AM-Net. As shown in Table 6 that the aging face generated by AM-Net achieves good results in the SSR-Net model test. AM-Net's facial aging on the generated test dataset is very close to the real aging face. It is shown that AM-Net can learn aging mechanisms at other ages and generate the aging faces that fit within the corresponding age range.

4.4.4. Aging-Net & Makeup-Net vs. Makeup-Net & Aging-Net

In this section, we will test how different learning sequences will affect the aging effect and makeup effect: (1) The network first learns the facial aging mechanisms, and then generates the results of facial makeup for different ages; (2) The network first generates faces with different makeup styles, and then utilizes these makeup faces to generate face aging images. In the evaluation of the aging effect, We still adopt the SSR-Net model [54] for age prediction and then calculate the accuracy to assess the effect of facial aging. Besides, we also ask volunteers to give corresponding scores for aging results. We ask volunteers to evaluate the generated images to obtain makeup results. We can learn from Table 4 that Method-2 performs well in both the aging task and the makeup task, which means that the result of applying makeup to the face first, and then aging the makeup face has little effect on facial aging and makeup. However, Method-1 has a significant impact on the effect of facial aging, which leads to a low evaluation of facial aging, but the makeup effect has received a high rating. That is to say, the AM-Net network first learns the facial aging mechanism, and then the result of applying makeup on the generated

face performs poorly on the task of facial aging, and performs well on facial makeup. It shows that the final makeup work has an effect on the facial aging task, but does not affect the makeup effect. As can be seen from Table 5, the effect is to make the generated face look younger. From Tables 4 and 5, we can conclude that with the increase of age, some changes will be found on their faces such as wrinkles and skin, but makeup can slightly conceal these facial skin tones, wrinkles, and other defects, thus making people look young.

4.4.5. The difference between PatchAutoEncoder and AutoEncoder

As shown in Fig. 4, PatchAutoEncoder follows the network structure of Encoder in AutoEncoder and modifies the network structure of Decoder. PatchAutoEncoder consists of two parts: Encoder that maps the input samples to the feature space and PatchDecoder that is used to synthesize the target image. PatchDecoder adds the calculation of texture loss on the basis of Decoder. While generating the target image, PatchDecoder extracts the features of each level in the image by calculating the texture loss in each level, so as to obtain the texture feature information of the original image. This process makes the generated images look more natural and real.

4.4.6. Analysis of AM-Net

In this part, we provide an overall analysis of AM-Net's network framework and performance. To solve the problem of facial makeup transfer at different ages, we construct the AM-Net model. In AM-Net, Aging-Net is composed of PathAutoEncoder and Estimator. PathAutoEncoder is used to capture the aging pattern of the character. The Estimator is used to distinguish the facial features at multiple scales, and the perceptual loss is applied externally to preserve the identity Information is not lost. In the past, many methods only performed aging operations to achieve the effect of facial aging. Although the faces generated by these methods appear realistic in terms of color, wrinkles, etc., they ignore what the face looks like in reality. For example, many faces in film or entertainment are beautified to look smooth and round. But most faces are somewhat rough, and many methods focus on aging while ignoring how the skin of the faces looks in reality, which makes the generated images unnatural in terms of aging, although very realistic. Different from the existing methods [33–35,56,57], we design texture loss to capture the texture features of the face and make the generated faces more natural. In terms of facial aging, the methods of the baseline model [34,49] have more or fewer problems such as unprotected identity, unrealistic images, and missing pixels. We can learn from Fig. 17 that the facial aging features generated by AM-Net are not limited to the corners of the mouth, face, and cheeks. AM-Net can also learn features such as the forehead, wrinkles at the corners of the eyes, and even hair color changes. Also, the facial skin generated by AM-Net appears more natural and real. Makeup-Net uses multiple losses include perception loss, confrontation loss, and cycle consistency loss to achieve the facial makeup transfer. To make the facial makeup effect more visible, we follow the idea of BeautyGAN [8] on the instance-level makeup transfer, adopt histogram matching (HM), and introduce additional histogram loss at the pixel level.

Through extensive experiments, we find that although faces such as wrinkles and skin will change slightly with increasing age, makeup can conceal these facial complexion, wrinkles and other defects. Even when people reach middle age, some of them still look young and beautiful after makeup. Besides, different makeup styles will also give people a different feeling. All in all, the effects of facial makeup for different ages vary from person to person and are also affected by makeup style. In daily life, even middle-aged people can make themselves look more beautiful through makeup.

5. Conclusion

In this paper, we propose a novel method of facial makeup transfer (AM-Net), which can well realize facial makeup transfer for different ages. AM-Net is composed of two sub-networks: Aging-Net and Makeup-Net. Aging-Net learns the aging mechanism of faces and then applies this aging mechanism to Makeup-Net. Then, Makeup-Net realizes facial makeup transfer. In the end, AM-Net realizes the facial makeup transfer for different ages. Extensive experiments show that our AM-Net can generate makeup images of different aging faces without losing the identity information. Furthermore, the experiments also prove the effectiveness of our proposed framework. What is more, AM-Net is a universal framework that can not only transfer facial makeup for different ages but also transfer facial makeup for different facial emotions, such as makeup from happy to angry or even angry. In future work, we will implement the current research results into specific applications to improve the quality of our lives.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, 2020, arXiv:1703.10593.
[2] Z. Yi, H. Zhang, P. Tan, M. Gong, DualGAN: Unsupervised dual learning for image-to-image translation, 2018, arXiv:1704.02510.
[3] T. Kim, M. Cha, H. Kim, J.K. Lee, J. Kim, Learning to discover cross-domain relations with generative adversarial networks, 2017, arXiv:1703.05192.
[4] D. Guo, T. Sim, Digital face makeup by example, 2009, pp. 73–79, http://dx.doi.org/10.1109/CVPR.2009.5206833.
[5] C. Li, K. Zhou, S. Lin, Simulating makeup through physics-based manipulation of intrinsic image layers, 2015, pp. 4621–4629, http://dx.doi.org/10.1109/CVPR.2015.7299093.
[6] W.-S. Tong, C.-K. Tang, M. Brown, Y.-Q. Xu, Example-based cosmetic transfer, 2007, pp. 211–218, http://dx.doi.org/10.1109/PG.2007.31.
[7] S. Liu, X. Ou, R. Qian, W. Wang, X. Cao, Makeup like a superstar: Deep localized makeup transfer network, 2016.
[8] T. Li, R. Qian, C. Dong, S. Liu, Q. Yan, W. Zhu, L. Lin, BeautyGAN: Instance-Level Facial Makeup Transfer with Deep Generative Adversarial Network, Association for Computing Machinery, 2018, pp. 645–653.
[9] Y. Fu, G. Guo, T. Huang, Age synthesis and estimation via faces: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 1955–1976, http://dx.doi.org/10.1109/TPAMI.2010.36.
[10] I. Kemelmacher, S. Suwajanakorn, S. Seitz, Illumination-aware age progression, in: Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014, http://dx.doi.org/10.1109/CVPR.2014.426.
[11] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, N. Sebe, Recurrent face aging, 2016, http://dx.doi.org/10.1109/CVPR.2016.261.
[12] B.-C. Chen, C.-S. Chen, W. Hsu, Cross-age reference coding for age-invariant face recognition and retrieval, 2014, pp. 768–783, http://dx.doi.org/10.1007/978-3-319-10599-4_49.
[13] G. Panis, A. Lanitis, An overview of research activities in facial age estimation using the FG-NET aging database, vol. 8926, 2015, pp. 737–750, http://dx.doi.org/10.1007/978-3-319-16181-5_56.
[14] K. Ricanek, T. Tesafaye, MORPH: A longitudinal image database of normal adult age-progression, in: FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, Vol. 2006, 2006, pp. 341–345, http://dx.doi.org/10.1109/FGR.2006.78.
[15] R. Rothe, R. Timofte, L. Van Gool, Deep expectation of real and apparent age from a single image without facial landmarks, Int. J. Comput. Vis. 126 (2018) http://dx.doi.org/10.1007/s11263-016-0940-3.
[16] A. Lanitis, Evaluating the performance of face-aging algorithms, in: 2008 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008, 2008, pp. 1–6, http://dx.doi.org/10.1109/AFGR.2008.4813349.
[17] X. Shu, J. Tang, H. Lai, L. Liu, S. Yan, Personalized age progression with aging dictionary, 2015, pp. 3970–3978, http://dx.doi.org/10.1109/ICCV.2015.452.
[18] J. Suo, S. Zhu, S. Shan, X. Chen, A compositional and dynamic model for face aging, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 385–401, http://dx.doi.org/10.1109/TPAMI.2009.39.

[19] A. Lanitis, C.J. Taylor, T.F. Cootes, Toward automatic simulation of aging effects on face images, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 442–455, http://dx.doi.org/10.1109/34.993553.

[20] N. Ramanathan, R. Chellappa, Modeling age progression in Young faces, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, CVPR'06, 2006, pp. 387–394, http://dx.doi.org/10.1109/CVPR.2006.187.

[21] Y. Tazoe, H. Gohara, A. Maejima, S. Morishima, Facial aging simulator considering geometry and patch-tiled texture, in: ACM SIGGRAPH 2012 Posters, in: SIGGRAPH '12, Association for Computing Machinery, New York, NY, USA, 2012, http://dx.doi.org/10.1145/2342896.2343002.

[22] I. Kemelmacher-Shlizerman, S. Suwajanakorn, S.M. Seitz, Illumination-aware age progression, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2014.

[23] B. Tiddeman, M. Burt, D. Perrett, Prototyping and transforming facial textures for perception research, IEEE Comput. Graph. Appl. 21 (5) (2001) 42–50, http://dx.doi.org/10.1109/38.946630.

[24] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, Vol. 27, Curran Associates, Inc., 2014, pp. 2672–2680.

[25] A. Dosovitskiy, T. Brox, Generating images with perceptual similarity metrics based on deep networks, 2016.

[26] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017.

[27] N. Sun, Unsupervised cross-view facial expression image generation and recognition, 2020.

[28] M. Mirza, S. Osindero, Conditional generative adversarial nets, 2014, CoRR, abs/1411.1784.

[29] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, 2016, arXiv:1511.06434.

[30] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training GANs, 2016, arXiv:1606.03498.

[31] Y. Wu, N. Thalmann, R. Dufour, D. Thalmann, A plastic-visco-elastic model for wrinkles in facial animation and skin aging, 2002.

[32] Y. Wang, Z. Zhang, W. Li, F. Jiang, Combining tensor space analysis and active appearance models for aging effect simulation on face images, IEEE Trans. Syst. Man Cybern. B 42 (4) (2012) 1107–1118.

[33] S. Liu, Y. Sun, D. Zhu, R. Bao, W. Wang, X. Shu, S. Yan, Face aging with contextual generative adversarial nets, 2018, arXiv:1802.00237.

[34] Z. Zhang, Y. Song, H. Qi, Age progression/regression by conditional adversarial autoencoder, 2017, arXiv:1702.08423.

[35] X. Tang, Z. Wang, W. Luo, S. Gao, Face aging with identity-preserved conditional generative adversarial networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7939–7947, http://dx.doi.org/10.1109/CVPR.2018.00828.

[36] M. Duan, K. Li, K. Li, Q. Tian, A novel multi-task tensor correlation neural network for facial attribute prediction, ACM Trans. Intell. Syst. Technol. (TIST) 12 (1) (2020) 1–22.

[37] M. Duan, K. Li, X. Liao, K. Li, Q. Tian, Features-enhanced multi-attribute estimation with convolutional tensor correlation fusion network, ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) 15 (3s) (2019) 1–23.

[38] X. Shu, J. Tang, H. Lai, L. Liu, S. Yan, Personalized age progression with aging dictionary, in: 2015 IEEE International Conference on Computer Vision, ICCV, 2015, pp. 3970–3978, http://dx.doi.org/10.1109/ICCV.2015.452.

[39] x. shu, J. Tang, Z. Li, H. Lai, L. Zhang, S. Yan, Personalized age progression with bi-level aging dictionary learning, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 905–917.

[40] Y. Sun, J. Tang, X. Shu, Z. Sun, M. Tistarelli, Facial age synthesis with label distribution-guided generative adversarial network, IEEE Trans. Inf. Forensics Secur. 15 (2020) 2679–2691, http://dx.doi.org/10.1109/TIFS.2020.2975921.

[41] S. Liu, Y. Sun, D. Zhu, R. Bao, W. Wang, X. Shu, S. Yan, Face aging with contextual generative adversarial nets, in: Proceedings of the 25th ACM International Conference on Multimedia, in: MM '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 82–90, http://dx.doi.org/10.1145/3123266.3123431.

[42] C. Shi, J. Zhang, Y. Yao, Y. Sun, H. Rao, X. Shu, CAN-GAN: Conditioned-attention normalized GAN for face age synthesis, Pattern Recognit. Lett. 138 (2020) 520–526.

[43] M. Duan, K. Li, Q. Liao, Q. Tian, DEF-Net: A face aging model by using different emotional learnings, IEEE Trans. Circuits Syst. Video Technol. (2021).

[44] M. Duan, A. Ouyang, G. Tan, Q. Tian, Age estimation using aging/rejuvenation features with device-edge synergy, IEEE Trans. Circuits Syst. Video Technol. 31 (2) (2020) 608–620.

[45] D. Mingxing, K. Li, L. Xie, Q. Tian, B. Xiao, Towards multiple black-boxes attack via adversarial example generation network, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 264–272.

[46] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, J. Choo, StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation, 2018, arXiv:1711.09020.

[47] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.

[48] L. Gatys, A.S. Ecker, M. Bethge, Texture synthesis using convolutional neural networks, in: C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol.28, Curran Associates, Inc., 2015, pp. 262–270, URL https://proceedings.neurips.cc/paper/2015/file/a5e00132373a7031000fd987a3c9f87b-Paper.pdf.

[49] H. Yang, D. Huang, Y. Wang, A.K. Jain, Learning face age progression: A pyramid architecture of GANs, 2017, CoRR, abs/1711.10352.

[50] B.-C. Chen, C.-S. Chen, W. Hsu, Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset, IEEE Trans. Multimed. 17 (2015) 1.

[51] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2015, arXiv:1512.03385.

[52] D. Guo, T. Sim, Digital face makeup by example, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 73–79, http://dx.doi.org/10.1109/CVPR.2009.5206833.

[53] S. Liu, X. Ou, R. Qian, W. Wang, X. Cao, Makeup like a superstar: Deep localized makeup transfer network, in: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, in: IJCAI'16, AAAI Press, 2016, pp. 2568–2575.

[54] T.-Y. Yang, Y.-H. Huang, Y.-Y. Lin, P.-C. Hsiu, Y.-Y. Chuang, SSR-Net: A compact soft stagewise regression network for age estimation, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, International Joint Conferences on Artificial Intelligence Organization, 2018, pp. 1078–1084, http://dx.doi.org/10.24963/ijcai.2018/150.

[55] M. Inc, Face++ research toolkit, http://www.faceplusplus.com.

[56] C.N. Duong, K. Luu, K.G. Quach, T.D. Bui, Longitudinal face modeling via temporal deep restricted Boltzmann machines, 2016, arXiv:1606.02254.

[57] Z. Qawaqneh, A. Abu Mallouh, B.D. Barkana, Deep convolutional neural network for age estimation based on VGG-face model, 2017, ArXiv E-Prints, arXiv:1709.01664.