

A³GAN: An Attribute-Aware Attentive Generative Adversarial Network for Face Aging

Yunfan Liu¹[✉], Qi Li¹[✉], Member, IEEE, Zhenan Sun¹[✉], Senior Member, IEEE, and Tieniu Tan, Fellow, IEEE

Abstract—Face aging has received significant research attention in recent years. Although great progress has been achieved with the success of Generative Adversarial Networks (GANs) in synthesizing realistic images, most existing GAN-based face aging methods have two main problems: 1) unnatural changes of high-level semantic information due to the insufficient consideration of prior knowledge of input faces, and 2) distortions of low-level image content (e.g., modifications in age-irrelevant regions). In this article, we introduce A³GAN, an Attribute-Aware Attentive face aging model to address the above issues. Facial attribute vectors are regarded as the conditional information and embedded into both the generator and discriminator, encouraging synthesized faces to be faithful to attributes of corresponding inputs. To improve the visual fidelity of generation results, we leverage the attention mechanism to restrict modifications to age-related areas and preserve image details. Unlike previous works with attention modules, we introduce face parsing maps to help the generator distinguish image regions of interest and suppress attention activation elsewhere. Moreover, the wavelet packet transform is employed to capture textural features at multiple scales in the frequency space. Extensive experimental results demonstrate the effectiveness of our model in synthesizing photo-realistic aged face images and achieving state-of-the-art performance on popular datasets.

Index Terms—Generative adversarial networks, face aging, facial attribute, attention mechanism, wavelet packet transform.

I. INTRODUCTION

FACE aging, also known as age progression or age synthesis, refers to rendering a given face image with realistic aging effects while still preserving personalized features [1]–[3]. Applications of face aging techniques range

Manuscript received September 9, 2020; revised December 24, 2020; accepted February 15, 2021. Date of publication March 18, 2021; date of current version April 1, 2021. This work was supported in part by the Natural Science Foundation of China under Grant U1836217, Grant 62076240, Grant 61721004, and Grant 61702513; in part by the Artificial Intelligence Research, Chinese Academy of Sciences (CAS-AIR); and in part by the Shandong Provincial Key Research and Development Program (Major Scientific and Technological Innovation Project) under Grant 2019JZZY010119. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. William R. Schwartz. (*Corresponding author: Zhenan Sun*)

Yunfan Liu, Zhenan Sun, and Tieniu Tan are with the National Laboratory of Pattern Recognition, Center for Research on Intelligent Perception and Computing, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: yunfan.liu@cripac.ia.ac.cn; znsun@nlpr.ia.ac.cn; tnt@nlpr.ia.ac.cn).

Qi Li is with the National Laboratory of Pattern Recognition, Center for Research on Intelligent Perception and Computing, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the Artificial Intelligence Research, Chinese Academy of Sciences, Qingdao 266300, China (e-mail: qli@nlpr.ia.ac.cn).

Digital Object Identifier 10.1109/TIFS.2021.3065499

from social security to digital entertainment, e.g. predicting the contemporary appearance of lost individuals or wanted suspects based on outdated photos, and improving the performance of face recognition systems in the cross-age verification scenario [4]–[8]. Because of the significant practical value, face aging has received considerable research attention but remains challenging due to its intrinsic complexity.

In the last two decades, face aging has witnessed impressive progress and numerous approaches have been proposed to address this problem. With the remarkable success of Generative Adversarial Networks (GANs) [9] in image synthesis and translation tasks, many recent studies resort to GAN-based frameworks to solve the face aging problem [3], [10]–[15]. These methods directly model the mapping function between distributions of face images at different ages, and translate test faces into the target age group via learned mappings with high visual fidelity.

Since multiple face images of the same subject at different ages are prohibitively expensive to collect in practice, most GAN-based methods resort to unpaired face aging data to train the model. However, these approaches mainly focus on simulating mappings between image contents while neglecting other critical semantic conditional information of the input (e.g., facial attributes), and thus fail to regulate the training process accordingly. Concretely, if no high-level conditional information is considered, a given young face image might map to multiple elderly face candidates in unpaired scenarios, which may mislead the model to establish translation patterns other than aging. Consequently, serious ghosting artifacts and even incorrect facial attributes may appear in synthesized face images, which seriously damages the authenticity and rationality of generation results. For example, Fig. 1 shows several face aging results with mismatched attributes. In the rightmost face aging result under ‘Gender’, beard is mistakenly attached to the input female face image, which is almost impossible to happen in the natural aging process. This is because the model learns that growing a beard is a typical sign of aging, but fails to recognize that this does not happen to a woman since no conditional information of the test face is involved in the training process.

In order to preserve personalized characteristics of input faces, many recent face aging studies attempt to supervise generation results by enforcing the identity consistency [3], [10]–[15]. However, as shown in Fig. 1, the identity of the test face is well preserved in the output for all sample results, nevertheless, unnatural changes of facial attributes

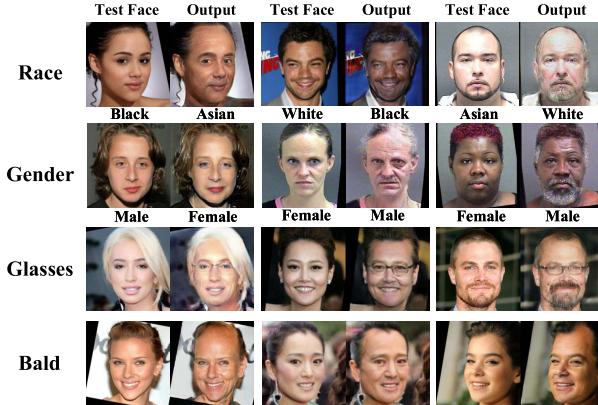


Fig. 1. Examples of face aging with mismatched facial attributes, which are generated by face aging model without facial attribute embedding. Four attributes (Race, Gender, Glasses, and Bald) are considered and three sample results are presented for each. Labels of ‘Race’ and ‘Gender’ are all obtained via the public face analysis API of Face++ [16] and placed under each image.

could still be observed. This suggests that well-maintained identity information does **NOT** imply reasonable aging results when training with unpaired data. Therefore, merely enforcing identity consistency is insufficient to eliminate matching ambiguities, and thus fails to achieve satisfactory face aging performance in unpaired training scenarios.

In addition to undesired changes of facial attributes, another critical problem of existing GAN-based face aging method is that image contents irrelevant to age progression (e.g., background or cloth) are not well preserved in the output. Basically, from the perspective of conditional image translation, face aging could be considered as adding representative signs of aging (e.g., wrinkles, eye bags, and laugh lines) to the input face image. Therefore, image modifications are supposed to be restricted to those regions highly relevant to age changes and image contents should be well preserved elsewhere. However, most existing GAN-based face aging methods do not enforce the constraint on regions of modification, instead, the pixel at each spatial location of the synthesis result is re-estimated by the generator.

To solve the above-mentioned issues, in this article, we propose A³GAN, a GAN-based framework for Attribute-Aware Attentive face aging. Different from existing methods in the literature, we involve semantic conditional information of the input by embedding facial attribute vectors into both the generator and discriminator, so that the model could be guided to output elderly face images with attributes faithful to the corresponding input. To improve the visual quality of synthesized face images, we leverage the attention mechanism to restrict modifications to age-related areas and preserve details in input images. Furthermore, based on the observation that signs of aging are mainly represented by local textures such as wrinkles, eye bags, and laugh lines, we employ wavelet packet transform in the critic network to extract features at multiple scales in the frequency space to enhance aging details.

Main contributions of this study are summarized as follows:

- An effective end-to-end GAN-based network, named A³GAN, is proposed to solve the face aging problem. Specifically, facial attributes are adopted to enforce

attribute consistency between input and generation results. Besides, a wavelet packet transform module is adopted to improve the quality of aging results.

- To improve the quality of synthesized images and suppress ghosting artifacts, attention mechanism is introduced to help restrict image modifications to age-related image regions. We further introduce face parsing maps to suppress attention activations in non-facial image regions.
- Extensive experiments are conducted to demonstrate the ability of the proposed method in rendering accurate aging effects and preserving information of both identity and facial attributes. Quantitative comparison with other seven advanced face aging benchmarks indicates that our method achieves state-of-the-art performance.

Compared to our previous work in [17], this article has the following extensions: 1) attention mechanism is introduced to help improve the visual quality of generation results via restricting modifications to image regions closely related to age progression; 2) generalization ability of the proposed method is investigated by testing the trained model on other widely used face datasets, including FG-NET [18] and CelebA [19]. 3) we refined our model to obtain better results, and a more thorough comparison with seven other GAN-based benchmark methods is provided to demonstrate the effectiveness of the proposed model in achieving reasonable and lifelike face aging results.

II. RELATED WORKS

A. Face Aging Synthesis

In the last few decades, a great number of algorithms have been proposed to solve the face aging problem, which could be divided into two categories: physical model-based methods and prototype-based methods.

Physical model-based methods are the initial explorations of face aging and they simulate changes of facial appearance w.r.t. time. As one of the earliest attempts, Todd *et al.* [20] model the profile growth via the revised cardioidal strain transformation. Subsequent works investigate the problem from various biological aspects including muscles and overall facial structures [21]–[25]. However, physical model-based algorithms are computationally expensive and difficult to generalize as they heavily depend on specific empirical aging rules.

As for data-driven prototyping approaches, Burt and Perrett [26] propose to divide faces into age groups, each represented by an average face, and regard differences between average faces as aging patterns. Following [26], many prototype-based methods are proposed to improve the face aging result [27]–[30]. However, averaging faces would blur the details of aging signs and damage the identity information in aging results.

With the rapid development of deep learning theory, deep generative models with temporal architectures are proposed to model age progression with hierarchically learned representations [31]–[33]. However, in most of these works, face image sequence over a long age span for each subject is required thus their potential in practical application is limited.

Recently, Generative Adversarial Networks (GANs) [9] have achieved remarkable success in generating visually plausible images, and many efforts have been made to solve the problem of face aging taking the advantage of adversarial learning [3], [10]–[12], [14], [15], [34]. Zhang *et al.* [10] propose a conditional adversarial autoencoder (CAAE) to achieve age progression and regression by traversing in a low-dimensional feature manifold. Li *et al.* [11] attend to three manually selected facial patches where age effects are likely to appear (i.e. forehead, eyes, and mouth), and separate generators are adopted to model the appearance change in these areas (GLCA-GAN). A wavelet-based decomposition and reconstruction module is then introduced in their following work [13] (WaveletGLCA-GAN) to enhance the quality of aging. By incorporating a conditional age vector, Wang *et al.* [12] achieve age progression to multiple target age groups with a single model. Using a pre-trained deep model in the discriminator network, Yang *et al.* [3] propose a GAN-based framework with pyramid-structured discriminator (PSD-GAN) to render aging effects. They further modified the discriminator to a parallel-structure and introduced an age vector as the conditional information in the following work [15]. He *et al.* [14] propose to share identity-related feature basis among all age translations, and all translations among all identities.

B. Attention Mechanism

Attention plays an important role in the human visual system as it serves as a high-level understanding of the scene and could guide the bottom-up processing of detailed objects [35]–[37]. In recent years, numerous attempts have been made to embed the attention mechanism into deep neural networks to improve the performance. Attention mechanism has been successfully applied to recurrent neural networks (RNN) and long short-term memory (LSTM) to tackle problems with sequential input, including neural machine translation [38], [39], visual question answering [40]–[42], and caption generation [43].

As for vision-related tasks, attention mechanism could be naturally introduced to guide the model to focus on specific image regions closely related to the target task. Wang *et al.* [44] propose a residual attention network which could generate attention-aware features for image classification. Woo *et al.* [45] explore the effectiveness of a light-weight general attention module, Convolutional Block Attention Module (CBAM), in improving the performance of deep models in various vision tasks. Pumarola *et al.* [46] adopt the spatial attention mechanism to synthesize face images with target expression. Attention is also widely used in solving image captioning [47] and saliency detection problems [48]–[52].

III. THE PROPOSED METHOD

A. Overview of the Framework

With an unpaired face aging dataset, a given young face image might map to many elderly face candidates during the training process, which may mislead the model into learning translations other than aging if no conditional information is

considered. To solve this problem, we present a GAN-based face aging model that takes both young face images and their semantic information (i.e. facial attributes) as input and outputs visually plausible aged faces with consistent facial attributes.

Our model mainly consists of two key components: an attribute-aware attentive generator G and a wavelet-based multi-pathway discriminator D . The generator G takes a young face image $I_y \in \mathbb{R}^{H \times W \times C}$ as input and predict the corresponding aged face I_o , while the discriminator D encourages generation results to be indistinguishable from generic face images. Unlike most existing face aging methods, the attribute of the input is considered as the conditional information and embedded into both G and D to ensure the attribute consistency. An overview of the proposed framework is shown in Fig. 2.

B. Attribute-Aware Attentive Generator

Most existing GAN-based face aging methods [3], [10]–[12], [14], [15] have two main problems:

- 1) Only images of young faces are taken as input to learn mappings between age groups, regardless of any prior knowledge that may have an influence on the visual pattern of age progression.
- 2) Although signs of aging concentrate on certain facial regions which only take up a small percentage of the entire image, pixel at each spatial location is re-estimated in the generation result, increasing the chance of introducing age-irrelevant changes and ghosting artifacts.

To solve these problems, in this article, we propose an attribute-aware attentive generator G to achieve fine-grained face aging. We employ an hourglass-shaped fully convolutional network as the backbone of the generator, which has achieved success in previous image translation studies [53], [54]. We also propose to incorporate both low-level image data (pixel values) and high-level semantic information (facial attributes) into the face aging model to regulate image translation patterns and reduce the ambiguity of image mappings. Concretely, the facial attribute is encoded in a binary vector $\alpha \in \mathbb{R}^{1 \times 1 \times N}$, where N is the number of total attributes to be preserved and each component indicates the status of one single attribute. Afterwards, α is replicated along the spatial dimension and then concatenated with the output of the last residual block, as they both contain high-level representations of the input image.

Considering that age progression is essentially the gradual emergence of aging signs, we naturally introduce the attention mechanism to guide the generator to concentrate more on image regions that aging signs are likely to appear. This is achieved by estimating an attention mask describing the contribution of each pixel in I_y to the final aging result. As shown in Fig. 2, the decoder network outputs two feature maps, an attention mask $M_A \in [0, 1]^{H \times W \times C}$ and an image map $M_I \in \mathbb{R}^{H \times W \times C}$. These two feature maps are then fed into a fusion layer along with the input image I_y to obtain the aged face image I_o , which could be formulated as,

$$I_o = M_A \odot I_y + (1 - M_A) \odot M_I \quad (1)$$

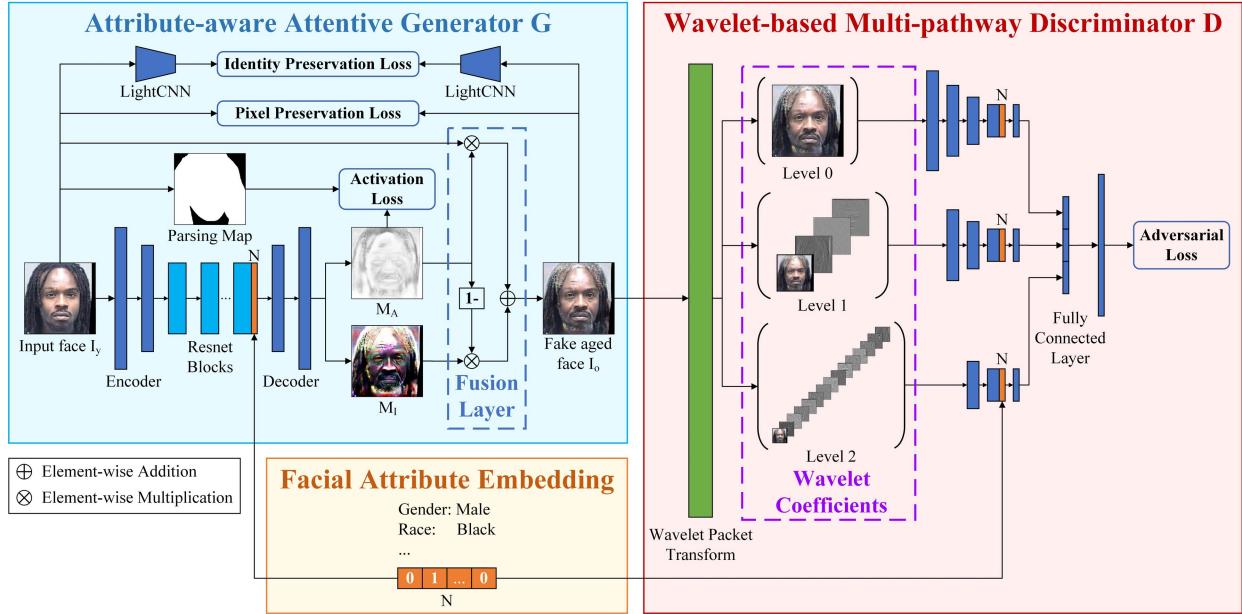


Fig. 2. An overview of the proposed A³GAN model. An hourglass-shaped generator G learns the age mapping and outputs lifelike elderly face images. A discriminator D is employed to distinguish synthesized face images from generic ones, based on multi-scale wavelet coefficients computed by the wavelet packet transform module. The N-dimensional binary attribute vector describing the input face image is embedded to both G and D to reduce matching ambiguity inherent to unpaired training data.

where \odot denotes element-wise product and M_A is replicated along the channel dimension to match the size of M_I and I_y . Note that M_A is a ‘soft’ attention map where the activation at each location lies in the interval [0,1] and is automatically optimized during the training process. Therefore, I_y and M_I could be merged seamlessly as in Eq. 1 without obvious sharp boundaries.

Intuitively, at each spatial location, the value of attention mask M_A indicates the proportion of the original input image being retained, in other words, to what extent the image map M_I contributes to the final output. For example, as shown in Fig. 2, brighter areas in M_A suggest those regions of the final output I_o retain more information from the input I_y , which are precisely image contents irrelevant to age changing (e.g. background and clothes). On the other hand, darker areas in M_A refer to regions in I_o that are more closely related to the image map M_I , that is, representative signs of face aging (i.e. forehead, hair, and laugh lines as shown in Fig. 2). The greatest advantage of adopting attention mechanism is that the generator could be guided to focus only on rendering age changing effects within specific regions, and pixels irrelevant to age progression could be directly obtained from the original input, resulting in more fine-grained image details with fewer ghosting artifacts.

C. Wavelet-Based Multi-Pathway Discriminator

In face aging tasks, a discriminator network D is introduced to distinguish synthetic aged face images from generic ones, and the generator learns to confuse D with outputs of high visual fidelity. In order to generate more accurate and lifelike aging details, Yang *et al.* [3] exploit a deep network with VGG-16 structure [55] pre-trained on an age classification task to extract age-related features conveyed by

faces. Although multi-scale representations could be obtained, storing and forwarding through a deep network would damage the efficiency of the model. Besides, pre-training also requires extra effort and might potentially limit the generalizability of the model due to the bias towards training dataset.

To overcome this issue, since typical signs of aging, e.g. wrinkles, laugh lines, and eye bags, could be regarded as local image textures, we adopt wavelet packet transform (WPT) to transform the input image to the frequency domain and capture textural features. For simplicity, fast wavelet transform is used to compute the coefficients of 2D Haar wavelet, and we perform wavelet coefficients decomposition by forwarding the input image through a convolutional layer consisting of pre-computed kernels with fixed weights. Specifically, multi-level WPT (see Fig. 3) is performed to provide a more comprehensive analysis of textures at multiple scales in the given image. Compared to extracting multi-scale features using a sequence of convolutional layers as in [3], the advantage of using WPT is that the computational cost is significantly reduced since wavelet coefficients could be calculated by simply forwarding through a single convolutional layer. Therefore, WPT greatly reduces the number of convolutions performed in each forwarding process. Although this part of the model has been simplified in terms of network structure, it still takes the advantage of multi-scale image texture analysis, which helps improve the visual fidelity of generated images.

D. Objective Functions and Training Procedure

The proposed method is supervised by a weighted sum of four losses: an *adversarial loss*, an *attention activation loss*, an *identity loss*, and a *pixel-level loss*.

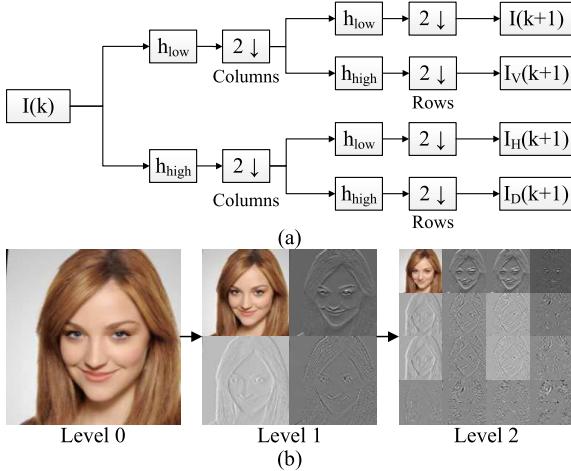


Fig. 3. Demonstration of wavelet packet transform. (a) Low-pass and high-pass decomposition filters (h_{low} and h_{high}) are applied iteratively to the input on the k -th level to compute wavelet coefficients on the next level; (b) a sample face image with its wavelet coefficients at different decomposing levels.

1) Adversarial Loss: The adversarial process between the generator G and discriminator D encourages synthetic results to be photo-realistic and indistinguishable from real data. Besides visual fidelity, attribute consistency is also guaranteed by involving the attribute of input face images as conditional information in the adversarial process.

To achieve these two goals, our discriminator network D is designed to take pair-wise input, i.e. aged face images and their corresponding attributes. Our goal is to make D gain the ability to discriminate generated aged face images from real ones and to tell whether the input face image contains the desired attribute. Therefore, to train the discriminator network, data pairs of real aged faces with attributes same as I_y , denoted by $\{I_o, \alpha\}$, are considered as positive samples. Negative samples include image pairs of synthesized aged faces $G(I_y, \alpha)$ and their attributes α , i.e. $\{G(I_y, \alpha), \alpha\}$, as well as images pairs of real aged faces and mismatched attributes, i.e. $\{I_o, \bar{\alpha}\}$.

Formally, the objective function for training the discriminative network D consists of two parts, that is, \mathcal{L}_{adv_att} for checking the attribute consistency and \mathcal{L}_{adv_auth} for image authenticity discrimination. Therefore, the adversarial loss could be formulated as,

$$\mathcal{L}_{adv_D} = \lambda_{att} \mathcal{L}_{adv_att} + \mathcal{L}_{adv_auth} \quad (2)$$

The parameter λ_{att} controls the relative importance between \mathcal{L}_{adv_att} and \mathcal{L}_{adv_auth} , which is initialized as 0 and then linearly increased during the training process. This enables D to firstly focus on discriminating fake images from real ones and then gradually adapts to the task of checking attribute consistency, which is critical for stabilizing the training process. We follow WGAN [56] and use the Wasserstein distance to measure the discrepancy between two data distributions. Therefore, \mathcal{L}_{adv_att} and \mathcal{L}_{adv_auth} are formulated as follows,

$$\begin{aligned} \mathcal{L}_{adv_att} &= -\mathbb{E}_{(I_o, \alpha) \sim P_o(I, \alpha)} [D(I_o, \alpha)] \\ &\quad + \mathbb{E}_{(I_o, \bar{\alpha}) \sim P_o(I, \bar{\alpha})} [D(I_o, \bar{\alpha})] \end{aligned} \quad (3)$$

$$\begin{aligned} \mathcal{L}_{adv_auth} &= -\mathbb{E}_{(I_o, \alpha) \sim P_o(I, \alpha)} [D(I_o, \alpha)] \\ &\quad + \mathbb{E}_{(I_y, \alpha) \sim P_y(I, \alpha)} [D(G(I_y, \alpha), \alpha)] \end{aligned} \quad (4)$$

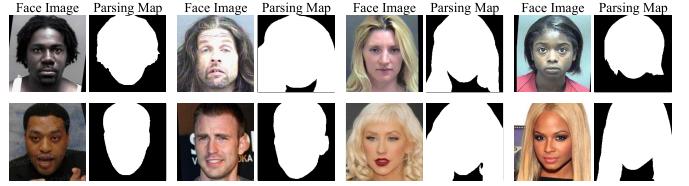


Fig. 4. Illustration of sample face parsing maps. Values of pixels in facial area and hair region (i.e. foreground) are set to 1, and pixels in other part of the image (i.e. background) are set to 0.

where P_y and P_o stand for the distribution of generic face images of young and old subjects, respectively. Notably, $P_o(I, \bar{\alpha})$ denotes the distribution of data pairs consisting real aged faces and mismatched attributes.

The generator network G is trained to confuse D with visually plausible synthetic images, and the objective function could be written as,

$$\mathcal{L}_{adv_G} = -\mathbb{E}_{(I_y, \alpha) \sim P_y(I, \alpha)} [D(G(I_y, \alpha), \alpha)] \quad (5)$$

Notably, since our model aims at rendering lifelike aging effects rather than transferring attributes of input face images, only correct attribute of young face images are fed into the generator in the training process.

2) Attention Activation Loss: Although the attention mechanism is adopted in the generator network to restrict image modifications, it does not necessarily mean that no activation will be assigned to irrelevant pixels.

To solve this problem, face parsing maps, which have long been used in understanding the semantic layout of images [57]–[60], are introduced to guide the generator to restrict attention activations within the foreground. Concretely, face parsing maps obtained by a pre-trained deep model are used to separate facial regions from background areas, that is, pixels in the foreground (i.e. face and hair) are set to 1, and 0 for those in the background. Examples of face parsing maps could be seen in Fig. 4. A straightforward way to apply parsing maps is simply multiplying them with M_A to filter out attention activations within background regions. However, this requires parsing maps of input images to be available at test time, which obviously limit the practical value of the trained generator.

Instead, we propose an attention activation loss to penalize the generator for total activations within the background region of attention maps. Mathematically, let P denote the face parsing map, the attention activation loss, denoted as L_{actv} , could be formulated as follows,

$$L_{actv} = \sum_{H,W} (1 - M_A) \cdot (1 - P) \quad (6)$$

where attention activations in the background region are summed up over the spatial dimension. Minimizing L_{actv} will encourage the generator to suppress attention activations in the background, and no parsing map is required at test time.

3) Identity Preservation Loss and Pixel-Level Loss: Although the goal of face aging is to modify a given face image to present aging effects, one key requirement is to preserve the identity-related information of the input. To this

end, we adopt the identity preserving loss to minimize the distance between input and output of the generator in the feature space embedding personalized characteristics. Specifically, we employ a pre-trained LightCNN model [61], denoted as ϕ_{id} , as the feature extractor and fix the parameters during the training process. To be concrete, the identity preserving loss is defined on the output of both the last pooling layer and the fully connected layer of ϕ_{id} , which could be formulated as,

$$\begin{aligned} \mathcal{L}_{id} = & \mathbb{E}_{(I_y, \alpha) \sim P_y(I, \alpha)} \left[\left\| \phi_{id}^{pool}(G(I_y, \alpha)) - \phi_{id}^{pool}(I_y) \right\|_F^2 \right] \\ & + \mathbb{E}_{(I_y, \alpha) \sim P_y(I, \alpha)} \left[\left\| \phi_{id}^{fc}(G(I_y, \alpha)) - \phi_{id}^{fc}(I_y) \right\|_2^2 \right] \end{aligned} \quad (7)$$

where ϕ_{id}^{pool} and ϕ_{id}^{fc} denote the output of the last pooling layer and the fully connected layer, respectively. Additionally, a pixel-level loss is also adopted to maintain the consistency of low-level image content between the input and output of the generator, which could be written as,

$$\mathcal{L}_{pix} = \mathbb{E}_{(I_y, \alpha) \sim P_y(I, \alpha)} \left[\|G(I_y, \alpha_y) - I_y\|_2^2 \right] \quad (8)$$

4) Overall Objective: To generate photo-realistic aged face with attributes faithful to the corresponding input, the overall objective function for optimizing the discriminator D could be formulated as

$$\min_{\|D\|_L \leq 1} \mathcal{L}_D = \mathcal{L}_{adv_D} = \lambda_{att} \mathcal{L}_{adv_att} + \mathcal{L}_{adv_auth} \quad (9)$$

where $\|D\|_L \leq 1$ denotes the 1-Lipschitz constraint [56] imposed on D , and is implemented by gradient penalty as proposed in WGAN-GP [62]. The objective for the generator G could be written as

$$\min_G \mathcal{L}_G = \mathcal{L}_{adv_G} + \lambda_{actv} \mathcal{L}_{actv} + \lambda_{id} \mathcal{L}_{id} + \lambda_{pix} \mathcal{L}_{pix} \quad (10)$$

where λ_{id} and λ_{pix} are hyperparameters for balancing the importance of \mathcal{L}_{id} and \mathcal{L}_{pix} w.r.t. the adversarial loss term, respectively. G and D are trained alternatively until reaching the optimality.

IV. EXPERIMENTS

In this section, we first introduce face aging datasets and then present implementation details of our model. After that, extensive qualitative and quantitative results are reported to demonstrate the effectiveness of the proposed method. Finally, ablation study is conducted to further explore the contribution of each component of our model.

A. Face Aging Datasets

Two publicly available face aging datasets, MORPH [63] and CACD [64], are employed in our experiments for both training and testing. **MORPH** contains 55,134 face images of 13,000 people, covering an age span of 16 to 77. Face images in MORPH capture near-frontal faces of collaborative subjects under uniform and moderate illumination with simple background. **CACD** contains 163,446 photos of 2,000 celebrities obtained in much less controlled (in-the-wild) conditions compared to MORPH. Consequently, large variations in terms

of pose, illumination, and expression (PIE variations) exist in CACD. Besides, due to the fact that images in CACD are collected via Google Image Search, there are mismatches between faces and associated labels provided (i.e. name and age), making it a very challenging dataset for accurate modeling of the face aging process.

Another two face datasets, FG-NET [18] and CelebA [19], are employed as test datasets to validate the generalization ability of the proposed model. **FG-NET** contains 1,002 face portraits of 82 subjects and is widely adopted in the test phase of previous works [3], [10], [13], [15], [28], [30]. **CelebA** is a large-scale face dataset featuring diverse facial attributes, which contains 202,599 face images with 40 attribute annotations for each sample. Similar to CACD, images in CelebA are also captured in the wild and cover large pose variations as well as background clutter.

B. Implementation Details

Image Normalization. Following the convention of previous studies [3], [10], [11], [13]–[15], [31], [33], we adopt the age span of 10 years for each age group and only consider adult aging as both MORPH and CACD do not contain images of children. In this way, faces are divided into four groups in terms of age, i.e. 30-, 31-40, 41-50, 51+, and only age translations from 30- to the other three age groups are considered. All face images in MORPH and CACD are aligned according to eye locations detected using MTCNN [65] and then cropped into size 256 × 256 × 3. As for FG-NET and CelebA, images are normalized according to facial landmarks provided along with the dataset.

For face parsing maps, we use a modified version of BiSeNet [66] pre-trained on CelebAMask-HQ dataset [67].¹ We convert the parsing maps to binary images with pixels within the foreground set to 1 (i.e. face and hair region), and 0 elsewhere (i.e. background, neck, and accessories).

Facial Attribute Labeling. As for facial attribute labeling, MORPH provides researchers with labels including age, gender, and race for each image. We choose ‘gender’ and ‘race’ to be the attributes required to be preserved, since these two attributes are relatively objective and guaranteed to remain unchanged during the natural aging process. For CACD, we go through the name list of celebrities and label corresponding images accordingly. This introduces noise in attribute labels due to the mismatching between annotated names and actual faces presented, which further increases the difficulty for our method to achieve a satisfying performance on this dataset. Since face images with race other than ‘white’ only take a small portion of the entire dataset, we only select ‘gender’ as the attribute to preserve. Facial attributes of FG-NET are detected via public face analysis APIs of Face++² [16], and for CelebA, we simply adopt facial attribute annotations provided along with the dataset. It is worthwhile to note that

¹<https://github.com/zllrunning/face-parsing.PyTorch>

²According to API documentations on the official website of Face++ (<https://console.faceplusplus.com>), the latest update of face APIs (Analyze API and Compare API) was in March, 2017. All quantitative results from Face++ API were obtained in August, 2019.

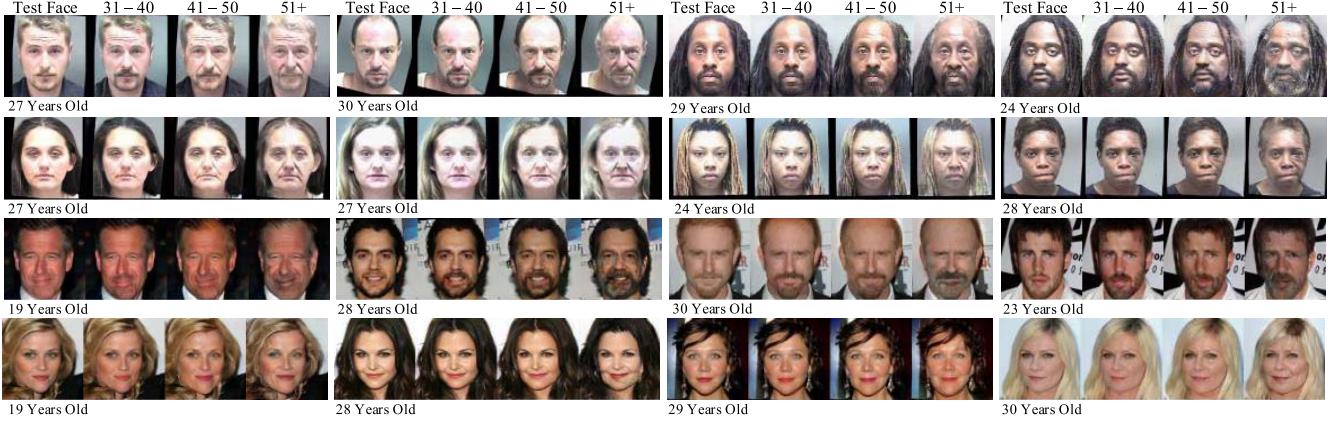


Fig. 5. Sample face aging results on MORPH (first two rows) and CACD (last two rows). The leftmost image of each result is the input test face image (age labeled below) and subsequent 3 images are synthesized elderly face images of the same subject in age group 31-40, 41-50 and 51+, respectively. Zoom in for a better view of aging details.

the proposed model is highly expandable, as researchers may choose whatever attributes to preserve by simply incorporating them in the conditional facial attribute vector.

Training Configurations. We choose Adam to be the optimizer of both G and D with learning rate $1e^{-4}$. As for trade-off parameters, λ_{att} , λ_{pix} and λ_{id} are set to 0.1, 10.0 and 0.02, respectively. The identity preserving loss is applied at every generator iteration, and the pixel-level loss is employed every 5 generator iterations, creating sufficient room for the generator to manipulate the input image. On both MORPH and CACD, the model is trained with batch-size of 16 for 30 epochs, and all experiments are conducted under 5-fold cross-validation.

C. Benchmark Methods

To demonstrate the effectiveness of our model, seven benchmark methods (CAAE [10], GLCA-GAN [11], WaveletGLCA-GAN [13], IPC-GAN [12], PSD-GAN [3], cPSD-GAN [15], and S2GAN [14]) are selected for comparison.

As for CAAE and IPC-GAN, code provided by corresponding authors^{3 4} are used for reproducing results, and hyper-parameters are fine-tuned to obtain the optimal results. Since GLCA-GAN requires fine-grained cropping of facial components, original experimental results are obtained from the authors and directly used for evaluation. Similarly, for WaveletGLCA-GAN, we report measurements computed on 400 samples on MORPH provided by the authors for attribute preservation, and refer to results in [13] for other evaluation metrics. As for PSD-GAN, we re-implemented the model and the VGG network in the discriminator is pre-trained on the same set of training samples as the entire GAN-based framework. For S2GAN and cPSD-GAN, we directly refer to results reported in their paper if possible for fair comparison, and evaluate their performance on attribute preservation based on samples provided by the authors. For S2GAN, aged results

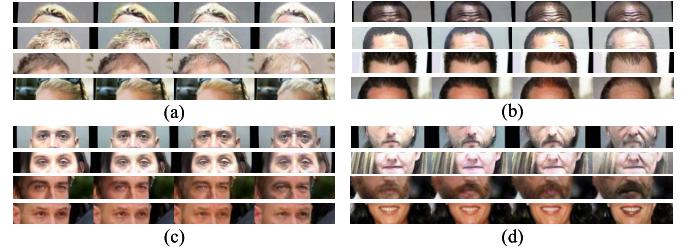


Fig. 6. Illustration of visual fidelity for different facial components: (a) Hair whitening; (b) Forehead wrinkle and receding hairline; (c) Eye region aging; (d) Mouth region aging. Zoom in for a better view of details.

of 45 and 147 test samples are obtained on MORPH and CACD from the author, respectively.

D. Qualitative Evaluation of A^3 GAN

1) Face Aging Results on MORPH and CACD: Sample face aging results on MORPH and CACD are shown in Fig. 5. Although input face images cover a wide range of gender, race, pose, and expression, the proposed method could generate visually appealing face aging results with coherent and diverse signs of age progression, including bald forehead, white hair, laugh lines, etc. Notably, compared to MORPH, generated faces in CACD present more fine-grained and subtler signs of aging, especially for female subjects. This observation reflects the difference in data distributions of MORPH and CACD, as CACD mainly contains face images of celebrities with apparent make-up, making them look generally younger than faces in MORPH. Quantitative results reported in Sec. IV-E.1 also confirms our conclusion. Closer inspection of aging details in different facial regions are presented in Fig. 6, where they are mainly located at hair, forehead, eyes, and mouth regions.

2) Demonstration of Spatial Attention Maps: The attention mechanism is employed in our model to restrict image modifications within age-related regions. Fig. 7 shows sample face aging results with corresponding attention maps. It could be observed that darker regions, which represent image

³CAAE: <https://github.com/ZZUTK/Face-Aging-CAAE>

⁴IPC-GAN: <https://github.com/dawei6875797/Face-Aging-with-Identity-Preserved-Conditional-Generative-Adversarial-Networks>

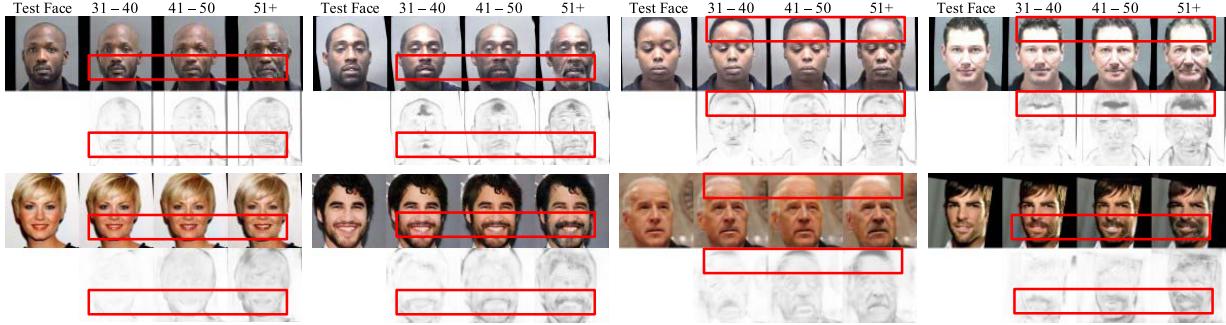


Fig. 7. Illustration of face aging results and corresponding attention maps on MORPH (first two rows) and CACD (second two rows). Darker regions suggest those areas of the face image receive more attention in the generation process, and brighter regions indicate that more information is retained from the original input image. Red boxes highlight image regions closely related to signs of aging.

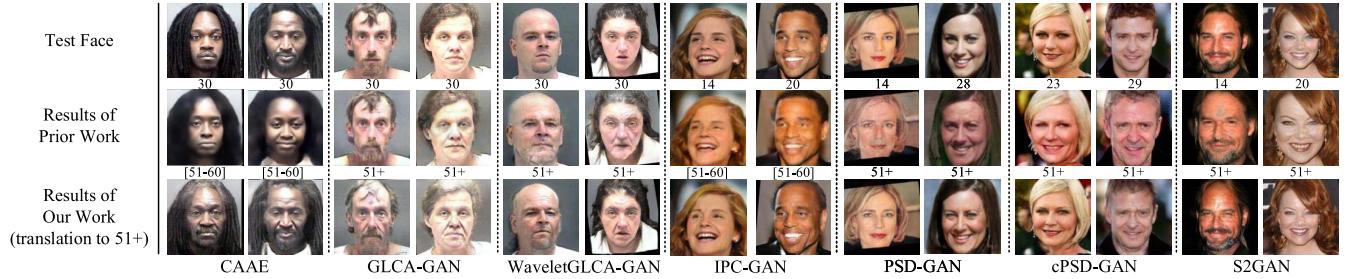


Fig. 8. Performance comparison with prior work on MORPH and CACD. Samples results of nine benchmark methods are presented in the second row with the target age (group) labeled below. Test face images and results obtained by the proposed model are shown in the first and last row, respectively. Zoom in for a better comparison of aging details.

areas receiving more attention in the generation process, are distributed mainly in facial areas closely related to signs of aging (e.g. hair, forehead, and mouth). Image content in these areas is modified to reflect age changes. On the other hand, pixels located in brighter regions are mainly retained from the original input image. This enables the generator to focus on synthesizing signs of aging, which is helpful for both preserving fine-grained textural details in the input image and improving the visual fidelity of generation results.

3) Comparison With Prior Work: To further demonstrate the effectiveness of our model, performance comparison is conducted between the proposed method and prior work (see Fig. 8).

As for CAAE, due to its incapability in jointly modeling age progression and identity translation, over-smoothed faces are generated and signs of aging could hardly be observed.

Although more obvious aging effects could be seen in results of GLCA-GAN and IPC-GAN, they are originally designed for face images of size 128×128 while our method works on higher resolution ($2\times$) with rich and enhanced details. In addition, GLCA-GAN adopts local generator networks to emphasize aging patterns in facial patches of forehead, eyes, and mouth, but it overlooks the importance of the hair region which is also critical in reflecting age changes. WaveletGLCA-GAN increases the resolution of input images to 256×256 , but it still suffers from ghosting artifacts introduced by fusing images patches synthesized by separate generators.

With the aid of multi-level face representations extracted by the pre-trained deep network in the discriminator, PSD-GAN

and cPSD-GAN are able to generate aged faces with high visual fidelity. However, they suffer from clear color distortions (tune of hair and skin) since the value of every single pixel is re-estimated by the generator rather than retained from the input image. Moreover, due to the lack of prior knowledge of input faces, masculine facial characteristics (e.g. stubble above the mouth) emerge in the aged faces of PDS-GAN, making the generation results look much less natural.

4) Cross-Dataset Validation: To evaluate the generalization ability of our model, cross-dataset validation experiments are conducted on FG-NET and CelebA with the model trained on CACD (see Fig. 9). Although input faces are sampled from different distributions, visually plausible aging results could still be obtained, demonstrating the effectiveness of our method in dealing with unseen face images. Notably, the activated regions in attention maps still concentrate on facial areas closely related to aging effects (e.g. white hair and laugh lines), which helps improve the visual quality of generation results by restricting the image region being modified.

E. Quantitative Evaluation

Apart from visual fidelity, the performance of the proposed model could also be quantitatively evaluated in the following aspects:

- **Aging Accuracy:** Synthesized aged faces are expected to present accurate aging signs that make them fall into the target age group.
- **Identity Preservation:** Personal characteristics of input faces are supposed to be preserved in generation results.

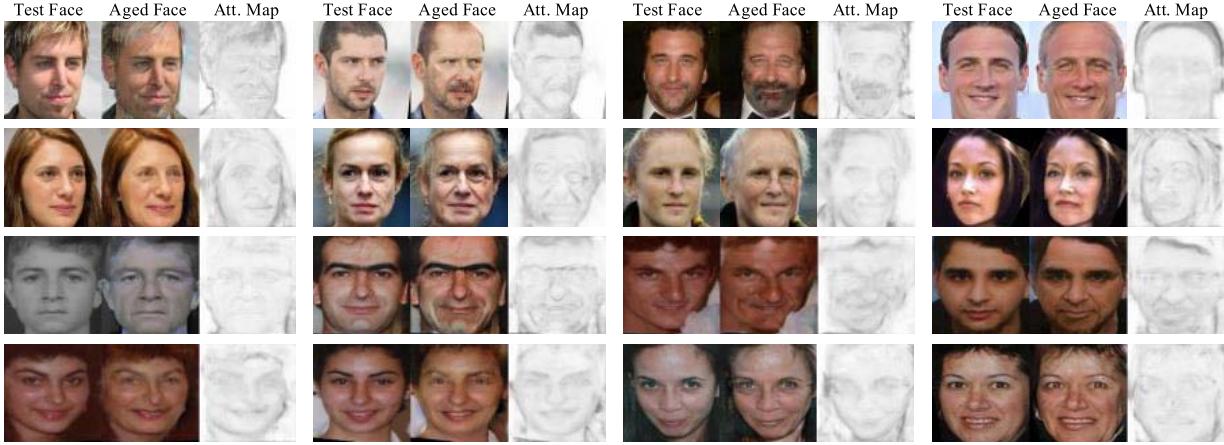


Fig. 9. Sample results achieved on CelebA (first two rows) and FG-NET (last two rows) with the model trained on CACD. For each sample, the first image is the test face and the image in the middle shows the aging result. The attention map (Att. Map) is shown on the right for each result. Zoom in for a better view of aging details.

TABLE I
RESULTS OF AGE ESTIMATION ON MORPH AND CACD

MORPH			CACD					
Age group	31 - 40	41 - 50	51 +	Age group	31 - 40	41 - 50	51 +	
Generic								
Generic	38.60 ± 7.43	47.74 ± 8.30	57.25 ± 8.29	Generic	38.50 ± 9.66	46.53 ± 10.69	53.41 ± 12.41	
Distributions of estimated ages (mean ± std.)								
CAAE	28.52 ± 5.26	32.25 ± 6.40	35.83 ± 7.49	CAAE	32.84 ± 7.67	36.21 ± 8.56	39.35 ± 9.65	
GLCA-GAN	43.79 ± 6.21	48.33 ± 6.85	53.25 ± 7.71	GLCA-GAN	38.15 ± 8.61	45.53 ± 9.04	53.30 ± 9.76	
WaveletGLCA-GAN	38.36 ± 6.51	46.90 ± 5.79	59.14 ± 6.00	WaveletGLCA-GAN	37.56 (-)	48.13 (-)	54.17 (-)	
IPC-GAN	36.45 ± 5.62	45.87 ± 5.80	55.63 ± 5.65	IPC-GAN	38.04 ± 8.59	47.75 ± 7.38	55.49 ± 8.71	
PSD-GAN	39.80 ± 6.90	50.09 ± 6.75	58.36 ± 6.91	PSD-GAN	42.97 ± 9.80	49.53 ± 9.23	55.75 ± 10.13	
cPSD-GAN	38.87 ± 7.52	48.03 ± 8.32	58.29 ± 8.76	cPSD-GAN	38.92 ± 9.73	46.95 ± 10.70	53.75 ± 12.45	
S2GAN	35.40 ± 2.80	45.20 ± 2.50	53.60 ± 2.70	S2GAN	36.00 ± 2.50	45.70 ± 2.20	55.30 ± 2.50	
Ours	38.84 ± 7.42	47.84 ± 7.03	56.68 ± 6.78	Ours	38.26 ± 9.36	47.25 ± 9.73	54.05 ± 9.17	
Difference of mean ages (against generic faces)								
CAAE	-10.08	-15.49	-21.42	CAAE	-5.66	-10.32	-14.06	
GLCA-GAN	+5.19	+0.59	-4.00	GLCA-GAN	-0.35	-1.00	-0.11	
WaveletGLCA-GAN	-0.24	-0.84	+1.89	WaveletGLCA-GAN	-0.94	+1.60	+0.76	
IPC-GAN	-2.15	-1.87	-1.62	IPC-GAN	-0.46	+1.22	+2.08	
PSD-GAN	+1.20	+2.35	+1.11	PSD-GAN	+4.47	+3.00	+2.34	
cPSD-GAN	+0.27	+0.29	+1.04	cPSD-GAN	+0.42	+0.42	+0.34	
S2GAN	-3.20	-2.54	-3.65	S2GAN	-2.50	-0.83	+1.89	
Ours	+0.24	+0.10	-0.57	Ours	-0.24	+0.72	+0.64	

- Attribute Consistency:** Facial attributes that should remain stable in the natural aging process, such as gender and race, are expected to be consistent between input and generated faces.

To evaluate the performance of the proposed method objectively, for both real and fake data, measurements of all metrics are conducted via the public face analysis API of Face++ [16]. For each evaluation metric, the results of seven other GAN-based frameworks (CAAE, GLCA-GAN, WaveletGLCA-GAN, IPC-GAN, PSD-GAN, cPSD-GAN, and S2GAN) are also reported for comparison.

1) *Aging Accuracy:* The age distribution of generated faces should match that of real faces from the same age group to represent accurate age simulation. In this experiment,

age distributions of both generic and synthetic faces are estimated and compared for all three target age groups (31-40, 41-50, 51+), and the results are shown in Table I. Clearly, mean estimated ages of synthesized faces show the trend of increasing age (38.84, 47.84, 56.68 on MORPH, and 38.26, 47.25, 54.05 on CACD), and are close to that of real faces (38.60, 47.74, 57.25 on MORPH, and 38.50, 46.53, 53.41 on CACD), demonstrating the effectiveness of our model in accurately simulating age progression with various time intervals.

Compared to our method, CAAE produces over-smoothed face images with subtle changes of appearance on both datasets, leading to insufficient facial changes and large errors in estimated ages. Although much more obvious aging

TABLE II

RESULTS OF FACE VERIFICATION ON MORPH AND CACD (THRESHOLD SET TO 73.395@FAR=1E-5 EXCEPT FOR S2GAN [14]). FOR S2GAN, ‘EASIEST’ REFERS TO AGE TRANSLATION BETWEEN (41-50, 51+) AND ‘HARDEST’ TO (30-, 51+)

Age group	MORPH			CACD			
	31 - 40	41 - 50	51 +	Age group	31 - 40	41 - 50	51 +
Face Verification Confidence							
CAAE	66.55 ± 9.99	65.20 ± 10.06	63.32 ± 10.17	CAAE	60.67 ± 10.65	59.30 ± 10.73	57.98 ± 10.82
GLCA-GAN	91.84 ± 2.33	90.42 ± 2.82	86.89 ± 4.20	GLCA-GAN	86.80 ± 5.65	85.30 ± 5.48	84.69 ± 5.47
WaveletGLCA-GAN	92.69 ± 1.85	91.14 ± 2.59	88.17 ± 3.48	WaveletGLCA-GAN	-	-	-
IPC-GAN	95.69 ± 0.45	93.92 ± 1.01	90.73 ± 2.35	IPC-GAN	91.86 ± 2.18	86.36 ± 4.86	87.85 ± 4.25
PSD-GAN	95.48 ± 0.80	92.64 ± 1.99	89.00 ± 3.29	PSD-GAN	94.82 ± 2.58	90.14 ± 5.53	90.38 ± 4.78
cPSD-GAN	94.65 ± 0.95	92.46 ± 1.87	88.12 ± 3.30	cPSD-GAN	94.61 ± 1.00	93.13 ± 1.68	91.22 ± 2.55
S2GAN	90.80 ± 1.93	95.75 ± 0.49	93.79 ± 1.20	S2GAN	90.50 ± 2.25	87.76 ± 3.59	90.55 ± 2.20
Ours	95.92 ± 0.66	92.76 ± 2.59	88.81 ± 3.92	Ours	96.19 ± 1.77	87.76 ± 3.20	90.55 ± 2.20
Face Verification Rate (%)							
CAAE	24.28	20.05	14.42	CAAE	9.20	7.04	5.10
GLCA-GAN	100.00	99.97	98.99	GLCA-GAN	96.09	95.79	95.29
WaveletGLCA-GAN	99.94	99.92	99.12	WaveletGLCA-GAN	97.71	96.07	95.25
IPC-GAN	100.00	100.00	99.48	IPC-GAN	100.00	97.95	97.36
PSD-GAN	100.00	100.00	99.42	PSD-GAN	99.83	97.67	98.50
cPSD-GAN	100.00	100.00	99.88	cPSD-GAN	100.00	100.00	99.98
S2GAN	Ave.: 99.69,	Easiest: 100.00 ,	Hardest: 96.08	S2GAN	Ave.: 98.91,	Easiest: 99.96,	Hardest: 94.08
Ours	100.00	100.00	99.53	Ours	99.94	99.61	98.85

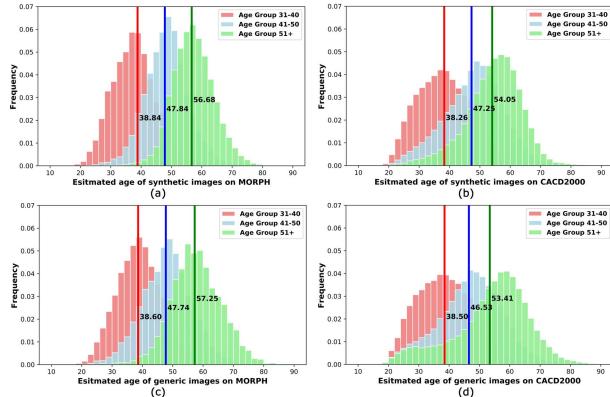


Fig. 10. Estimated age distributions of (a) synthetic faces on MORPH; (b) synthetic faces on CACD; (c) generic faces on MORPH; (d) generic faces on CACD.

signs could be synthesized by GLCA-GAN, on MORPH, stitching outputs of several local aging networks introduce additional ghosting artifacts in generation results of age group 31-40, causing large errors (+5.19) in estimated ages. WaveletGLCA-GAN improves the performance by introducing a wavelet decomposition and reconstruction module to retain textural information in the frequency domain. With the aid of a pre-trained VGG network adopted in the discriminator, PSD-GAN could effectively extract multi-level age-related representations and generate faces with clear aging signs. However, due to the matching ambiguity of facial attributes, translations in gender (female to male) take place when synthesizing aged faces, causing the overall age of generation results to be higher than generic faces, especially on CACD.

Further comparisons between detailed age distributions between real and generated face images are shown in Fig. 10. For each age group, it could be observed that the age

distribution on MORPH is more concentrated than that on CACD by comparing Fig. 10 (c) and Fig. 10 (d). This is because faces in CACD are obtained online via image search engine thus contain noisy age labels. By comparing Fig. 10 (a) and Fig. 10 (c) as well as Fig. 10 (b) and Fig. 10 (d), it is clear that age distributions of faces generated by our model well match that of real faces, indicating the effectiveness of the proposed method in rendering accurate age translations.

2) *Identity Preservation*: Face verification experiments are carried out to measure the similarity between real faces from age group 30- and their age-progressed counterparts in other age groups, and the results are reported in Table II.

On MORPH, our model achieves verification rates of 100.00%, 100.00%, and 99.53% on translation to age group 31-40, 41-50, and 51+, respectively. Although there are larger variations in pose, expression, and background textures in images of CACD, face verification rates of 99.94%, 99.61%, and 98.85% are obtained on three age groups, demonstrating the effectiveness of the proposed method in preserving identity information. Notably, as the time interval of age progression increases, both confidence scores and verification rates gradually decrease. This is reasonable since a larger age gap is reflected in more obvious aging signs (e.g. deeper wrinkles and eye bags), which may lower the similarity between faces from different age groups.

As for benchmark methods, personalized facial features are failed to be preserved in heavily blurred aging results generated by CAAE, causing poor face verification performance. To preserve as many facial details in the input face image as much, a residual connection is adopted in GLCA-GAN and WaveletGLCA-GAN by adding the input face image to the output of the generator. However, separate translations of different facial components inevitably introduce extra distortions and ghosting artifacts, which lead to a slightly

TABLE III
PRESERVATION RATE OF FACIAL ATTRIBUTES ON MORPH AND CACD

Age group	MORPH						CACD		
	Gender (%)			Race (%)			Gender (%)		
	31 - 40	41 - 50	51 +	31 - 40	41 - 50	51 +	31 - 40	41 - 50	51 +
CAAE	51.38	47.07	54.24	95.45	95.23	92.37	87.43	86.53	85.25
GLCA-GAN	96.44	95.90	94.85	93.69	91.79	91.48	95.46	95.51	94.65
WaveletGLCA-GAN*	72.50	69.62	83.38	83.88	83.12	83.62	-	-	-
IPC-GAN	96.87	97.45	96.75	97.11	96.88	90.57	94.79	90.18	93.24
PSD-GAN	96.62	95.94	93.28	96.61	91.77	91.42	87.56	83.19	75.72
cPSD-GAN	-	-	-	-	-	-	-	-	-
S2GAN†	93.33	88.89	86.67	100.00	100.00	100.00	74.83	71.43	65.99
Ours	97.41	97.58	96.92	97.68	96.36	93.28	99.00	98.59	98.00

* Results for WaveletGLCA-GAN on MORPH are computed based on 400 samples obtained directly from the authors.

† Results for S2GAN on MORPH and CACD are computed based on 45 and 147 samples obtained directly from the authors, respectively.

lower verification rate on generated face images of 51+. The performance of PSD-GAN on identity preservation is very close to ours, and the verification rate between 30- and 51+ on MORPH (99.42%) is clearly higher than reported in [3] (93.09%), indicating the quality of our re-implementation. Multiple parallel discriminators help cPSD-GAN gain better identity preservation performance on both datasets. S2GAN achieves lower face verification rate in cases with large age gaps (30- to 51+), since identity preservation loss is enforced only for input and output with the same age label [14].

3) *Facial Attribute Preservation*: In this experiment, the performance of facial attribute preservation is evaluated by comparing attributes of generic and synthetic faces estimated by the Face++ API, and results are shown in Table III. On MORPH, our model achieves preservation rate of 97.41%, 97.58%, 96.92% on ‘Gender’ and 97.68%, 96.36%, 93.28% on ‘Race’, for age mappings from 30- to 31-40, 41-50, and 51+, respectively. As for CACD, 99.00%, 98.59%, and 98.00% of generated faces in age group 31-40, 41-50, and 51+ have facial attributes consistent with the corresponding input, respectively.

According to Table III, it could be observed that the proposed method consistently outperforms other benchmarks by a clear margin in all cases, demonstrating the effectiveness of our model in preserving facial attributes beyond identity information during the face aging process. Specifically, gender characteristics are lost along with identity information in faces generated by CAAE, causing large errors in preservation rate on ‘Gender’. Among all four benchmarks with attribute preservation rate available, GLCA-GAN and IPC-GAN give better performance on maintaining facial attribute consistency. This is because they are applied to face images of lower resolution (128×128) with less textural details, which reduces the chance of introducing distortions of fine-grained image content. Although PSD-GAN achieves good results on aging accuracy and face verification, it suffers from inconsistent facial attributes between input and generated faces, due to the lack of prior knowledge regarding the input image.

F. Ablation Study

In this subsection, experiments are carried out to comprehensively analyze the contribution of each component

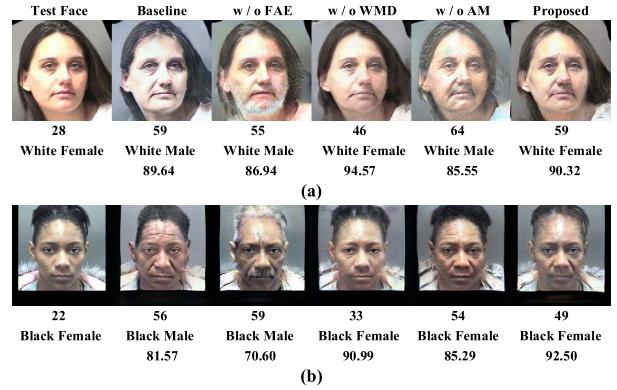


Fig. 11. Illustration of face aging results generated by different variants of the proposed model. For each subject, estimated ages (first row), facial attributes (second row), and confidence scores of face verification (third row) are reported. All quantitative results are obtained using Face++ API.

of the proposed model, namely, facial attribute embedding (FAE), wavelet-based multi-pathway discriminator (WMD), and attention mechanism (AM). We exclude each one of these modules (‘w/o FAE’, ‘w/o WMD’, and ‘w/o AM’) and study the impact using the same metrics as in previous sections (visual fidelity, aging accuracy, identity verification, and facial attribute preservation). Concretely, in ‘w/o FAE’, we remove network structures related to the facial attribute embedding mechanism as well as associated loss terms supervising the attribute consistency; in ‘w/o WMD’, we replace the discriminator with a variant similar to that in PSD-GAN [15], where multi-scale feature maps are extracted by a VGG-16 network pre-trained on the age estimation task instead of the WPT layer; in ‘w/o AM’, we remove M_I , M_A , and network layers associated with AM, and thus the final generation result is directly produced by the decoder in G .

Generation results obtained by different variants of the proposed model are shown in Fig. 11. It could be observed that aged faces synthesized via the baseline model have severe ghosting artifacts (e.g. hair area) and color distortion (e.g. facial skin). Since no semantic prior knowledge of the input face is considered in the aging process, masculine facial characteristics emerge, and gender reversal takes place.

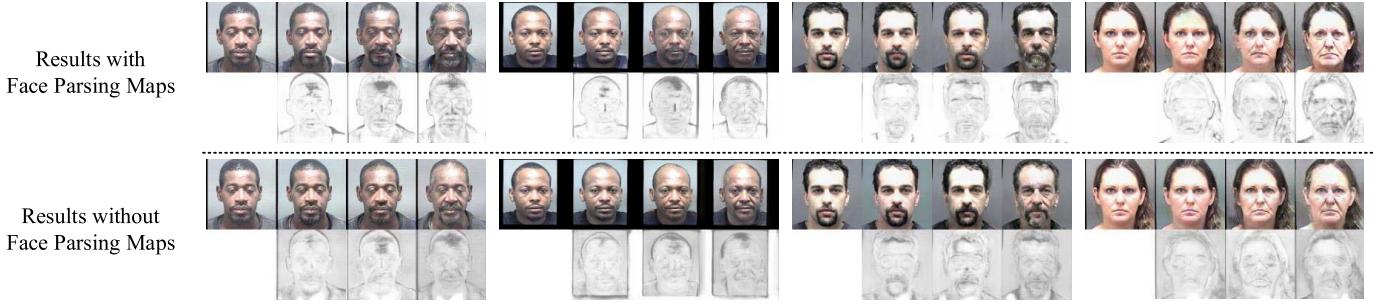


Fig. 12. Illustration of the effect of using face parsing maps to suppress attention activations in image regions unrelated to age changing. Zoom in for a better view of details.

TABLE IV
PERFORMANCE COMPARISON ON AGING ACCURACY BETWEEN VARIANTS OF THE PROPOSED MODEL (DIFFERENCE OF MEAN AGES AGAINST GENERIC FACES ARE SHOWN IN BRACKETS)

Age group	31-40	41-50	51+
Generic	38.60 ± 7.43	47.74 ± 8.30	57.25 ± 8.29
Baseline	$36.67 \pm 7.62 (-2.17)$	$45.21 \pm 7.98 (-2.53)$	$55.47 \pm 8.75 (-1.78)$
w/o FAE	$37.32 \pm 7.56 (-1.28)$	$46.88 \pm 7.68 (-0.86)$	$55.53 \pm 8.31 (-1.72)$
w/o WMD	$36.22 \pm 6.48 (-2.38)$	$46.47 \pm 7.83 (-1.27)$	$53.53 \pm 8.30 (-3.72)$
w/o AM	$37.23 \pm 7.12 (-1.37)$	$48.18 \pm 7.55 (+0.44)$	$54.84 \pm 8.03 (-2.41)$
Proposed	$38.84 \pm 7.42 (+0.24)$	$47.84 \pm 7.03 (+0.10)$	$56.68 \pm 6.78 (-0.57)$

Similarly, results obtained under the setting ‘w/o FAE’ also suffer from unnatural translations of facial attributes due to the lack of conditional information. However, it could be noticed that involving WMD and AM helps to capture more representative age-related facial features (e.g. beard in Fig. 11 (a), white hair and mustache in Fig. 11 (b)) and preserving image content in the input, respectively. Notably, although aged faces generated under setting ‘baseline’ and ‘w/o FAE’ have limitations in maintaining the consistency of facial attributes, this is not clearly reflected in results of age estimation or face verification shown in Table IV and Table V. Therefore, it could be concluded that merely enforcing identity consistency is insufficient in synthesizing aged faces reasonable in terms of facial attributes.

After adopting FAE, the unnatural translation of gender characteristics is greatly suppressed, as could be observed from results under ‘w/o WMD’ and ‘w/o AM’ in Fig. 11. However, closer inspection would clearly reveal that replacing WMD with ordinary PatchGAN discriminator (‘w/o WMD’) damages the performance of the model in capturing age-related texture details, resulting in relatively larger errors in estimated ages. As for results obtained under ‘w/o AM’, distortions of color and image content (e.g. facial contour and textual details of the hair area in Fig. 11 (b)) as well as ghosting artifacts (e.g. mouth and hair region in Fig. 11 (a)) could be observed, leading to lower verification confidence and even incorrect facial attribute recognition results. This confirms the contribution of the attention mechanism, that is, improving the visual fidelity of generation results by only attending to specific image regions closely related to age progression, and retaining textual details from the input face for the rest image areas.

The effect of face parsing map is shown in Fig. 12. Although attention activations mainly concentrate on image areas related to aging (e.g. forehead, mouth, and hair region), pixels in other part of the image still contribute to the final output. Consequently, there are minor changes in irrelevant regions (e.g. color of the background) and thus damage the quality of generation results. These undesired attention activations are clearly suppressed after involving face parsing maps in training, which improve the visual fidelity of age faces.

Quantitative results for ablation study are reported in Table IV and Table V. According to results in Table V, removing the facial attribute embedding component (‘w/o FAE’ and ‘Baseline’) would cause obvious performance drop in preservation rate for both ‘gender’ and ‘race’. From another perspective of view, while face verification rates have already reached a high level (over 98.00% for most cases), there is still relatively large room for improvement on preservation rates of facial attributes. Therefore, we could conclude that facial attribute consistency is complementary to identity permanence in rendering natural and reasonable face aging results with high visual fidelity.

In addition, from results in Table IV, it could be concluded that adopting wavelet-based multi-pathway discriminator (WMD) reduces the gap between age distributions of real and synthesized faces for all age mappings. This demonstrates the ability of WMD in capturing discriminative representations of age progression which is helpful in rendering more accurate signs of aging. Moreover, introducing the attention mechanism helps comprehensively improve the performance of the model on all experiments, indicating its effectiveness in generating aged faces with high visual quality.

TABLE V

PERFORMANCE COMPARISON ON FACIAL ATTRIBUTE PRESERVATION AND FACE VERIFICATION BETWEEN VARIANTS OF THE PROPOSED MODEL

Age group	Gender Pre. Rate (%)			Race Pre. Rate (%)			Face Veri. Rate (%)			Face Veri. Score		
	31-40	41-50	51+	31-40	41-50	51+	31-40	41-50	51+	31-40	41-50	51+
Baseline	97.05	95.35	92.20	97.04	94.85	91.18	100.00	99.99	97.66	95.75 ± 0.72	93.09 ± 2.31	86.23 ± 4.77
w/o FAE	96.95	96.11	92.93	95.20	95.84	88.32	100.00	99.99	97.68	95.31 ± 1.29	93.26 ± 2.20	86.56 ± 5.20
w/o WMD	96.84	96.93	96.14	97.44	96.74	91.66	100.00	99.96	97.02	95.68 ± 1.03	93.24 ± 2.34	88.41 ± 3.99
w/o AM	96.90	96.27	94.95	97.69	95.89	91.57	100.00	99.93	98.44	95.65 ± 0.90	92.17 ± 2.57	86.31 ± 4.47
Proposed	97.41	97.58	96.92	97.68	96.36	93.28	100.00	100.00	99.63	95.92 ± 0.66	92.76 ± 2.59	88.81 ± 3.92



Fig. 13. Sample failure cases on MORPH. Two typical categories of distortion could be observed, i.e. (a) face color distortion and (b) hair shape distortion.

G. Failure Cases

Although satisfactory results could be obtained by our method, the discussion on failure cases is still necessary to reveal shortcomings of the proposed framework and explore possible solutions. We have reviewed generation results of our method and found two typical categories of failure cases, i.e. **face color distortion** and **hair shape distortion**, as shown in Fig. 13. Concretely, face color distortion is often observed in aging results of white people, where the facial skin has become more red and ruddy compared to input faces. As for hair shape distortion, it is more frequently observed in aging results of black people with braids, where the hair region often fades into the background.

Although these two kinds of failure cases seem different in the image domain, they are caused by the same reason, that is, **the imbalance of data distributions between young and old training face images**. Closer inspection of the training dataset reveals that 1) the facial skin of aged white people tends to look redder compared to young faces, and 2) almost all aged black people in the training set have short hair and braids could hardly be observed. These two characteristics of aged faces are learned by the generator and reflected in the editing results, causing undesired image modifications and ghosting artifacts. Moreover, this could be regarded as the presentation of domain gap in the image level, while our work proposes to solve the domain gap in another level, i.e. the semantic attribute level. Therefore, a potential direction for future work

would be introducing prior knowledge to bridge the domain gap in image level, e.g. face parsing maps.

V. CONCLUSION AND FUTURE WORK

In this article, an attribute-aware attentive face aging model, named as A³GAN, is proposed to overcome two major limitations of existing face aging methods, i.e. unnatural translations of facial attributes and modifications to image contents irrelevant to age progression. Specifically, facial attributes of input images are considered as conditional information and embedded to both the generator and discriminator to encourage attribute consistency. Besides, the attention mechanism is adopted to restrict modifications to age-related regions and preserve image details from the input image for the rest area, improving the visual fidelity of generation results. Moreover, a wavelet packet transform module is employed to extract textual features, and a multi-pathway discriminator is designed to capture age-related representations in multiple scales. Extensive experimental results on MORPH and CACD demonstrate the effectiveness of the proposed model in rendering accurate aging effects while maintaining identity permanence and facial attribute consistency.

Although the proposed method achieves state-of-the-art performance in various experiments, it indeed has some limitations. Since existing face aging datasets are heavily biased towards White and Black people, aging patterns of other ethnic groups (e.g. Asian and Hispanic) receive much less attention. Besides, child aging is not investigated in this work due to insufficient data of child faces at different ages. Considering the above issues, collecting large-scale high-quality face images covering various ethnic and age groups could be one working direction in the future.

REFERENCES

- [1] N. Ramanathan, R. Chellappa, and S. Biswas, “Age progression in human faces: A survey,” *J. Vis. Lang. Comput.*, vol. 15, pp. 3349–3361, Apr. 2009.
- [2] Y. Fu, G. Guo, and T. S. Huang, “Age synthesis and estimation via faces: A survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1955–1976, Nov. 2010.
- [3] H. Yang, D. Huang, Y. Wang, and A. K. Jain, “Learning face age progression: A pyramid architecture of GANs,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 31–39.
- [4] J. Zhao *et al.*, “Towards pose invariant face recognition in the wild,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2207–2216.
- [5] J. Zhao *et al.*, “3D-aided deep pose-invariant face recognition,” in *Proc. 27th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2018, pp. 1184–1190.
- [6] J. Zhao *et al.*, “Look across elapse: Disentangled representation learning and photorealistic cross-age face synthesis for age-invariant face recognition,” in *Proc. 33rd AAAI Conf. Artif. Intell.*, vol. 33, Aug. 2019, pp. 9251–9258.

- [7] J. Zhao *et al.*, "Multi-prototype networks for unconstrained set-based face recognition," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 4397–4403.
- [8] J. Zhao, J. Xing, L. Xiong, S. Yan, and J. Feng, "Recognizing profile faces by imagining frontal view," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 460–478, Feb. 2020.
- [9] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2014, pp. 2672–2680.
- [10] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4352–4360.
- [11] P. Li, Y. Hu, Q. Li, R. He, and Z. Sun, "Global and local consistent age generative adversarial networks," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 1073–1078.
- [12] X. Tang, Z. Wang, W. Luo, and S. Gao, "Face aging with identity-preserved conditional generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7939–7947.
- [13] P. Li, Y. Hu, R. He, and Z. Sun, "Global and local consistent wavelet-domain age synthesis," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 11, pp. 2943–2957, Nov. 2019.
- [14] Z. He, M. Kan, S. Shan, and X. Chen, "S2GAN: Share aging factors across ages and share aging trends among individuals," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9440–9449.
- [15] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning continuous face age progression: A pyramid of GANs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 499–515, Feb. 2021.
- [16] Megvii Inc. *Face++ Research Toolkit*. Accessed: Jun. 12, 2020. [Online]. Available: <http://www.faceplusplus.com>
- [17] Y. Liu, Q. Li, and Z. Sun, "Attribute-aware face aging with wavelet-based generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, p. 11.
- [18] A. Lanitis, "Comparative evaluation of automatic age progression methodologies," *EURASIP J. Adv. Signal Process.*, vol. 8, no. 2, pp. 1–10, 2008.
- [19] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3730–3738.
- [20] J. T. Todd, L. S. Mark, R. E. Shaw, and J. B. Pittenger, "The perception of human growth," *Sci. Amer.*, vol. 242, no. 2, pp. 132–145, 1980.
- [21] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, Apr. 2002.
- [22] A. Golovinskiy, W. Matusik, H. Pfister, S. Rusinkiewicz, and T. Funkhouser, "A statistical model for synthesis of detailed facial geometry," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 1025–1034, Jul. 2006.
- [23] N. Ramanathan and R. Chellappa, "Modeling shape and textural variations in aging faces," in *Proc. 8th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Sep. 2008, pp. 1–8.
- [24] J. Suo, S.-C. Zhu, S. Shan, and X. Chen, "A compositional and dynamic model for face aging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 385–401, Mar. 2010.
- [25] Y. Tazoe, H. Gohara, A. Maejima, and S. Morishima, "Facial aging simulator considering geometry and patch-tiled texture," in *Proc. ACM SIGGRAPH Posters*, 2012, p. 1.
- [26] D. M. Burt and D. I. Perrett, "Perception of age in adult caucasian male faces: Computer graphic manipulation of shape and colour information," *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 259, no. 1355, pp. 137–143, 1995.
- [27] B. Tiddeman, M. Burt, and D. Perrett, "Prototyping and transforming facial textures for perception research," *IEEE Comput. Graph. Appl.*, vol. 21, no. 4, pp. 42–50, Sep. 2001.
- [28] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. M. Seitz, "Illumination-aware age progression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3334–3341.
- [29] X. Shu, J. Tang, H. Lai, Z. Niu, and S. Yan, "Kinship-guided age progression," *Pattern Recognit.*, vol. 59, pp. 156–167, Nov. 2016.
- [30] X. Shu, J. Tang, Z. Li, H. Lai, L. Zhang, and S. Yan, "Personalized age progression with bi-level aging dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 905–917, Apr. 2018.
- [31] W. Wang *et al.*, "Recurrent face aging," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2378–2386.
- [32] C. N. Duong, K. Luu, K. G. Quach, and T. D. Bui, "Longitudinal face modeling via temporal deep restricted Boltzmann machines," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5772–5780.
- [33] C. N. Duong, K. G. Quach, K. Luu, T. H. N. Le, and M. Savvides, "Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3755–3763.
- [34] J. Zhao *et al.*, "Dual-agent GANs for photorealistic and identity preserving profile face synthesis," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017, pp. 66–76.
- [35] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
- [36] R. A. Rensink, "The dynamic representation of scenes," *Vis. Cognition*, vol. 7, nos. 1–3, pp. 17–42, Jan. 2000.
- [37] M. Corbetta and G. L. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature Rev. Neurosci.*, vol. 3, no. 3, pp. 201–215, Mar. 2002.
- [38] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–15.
- [39] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 5998–6008.
- [40] H. Xu and K. Saenko, "Ask, attend and answer: Exploring question-guided spatial attention for visual question answering," *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 451–466.
- [41] Z. Yang, X. He, J. Gao, L. Deng, and A. Smola, "Stacked attention networks for image question answering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 21–29.
- [42] D. Yu, J. Fu, T. Mei, and Y. Rui, "Multi-level attention networks for visual question answering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4709–4717.
- [43] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 2048–2057.
- [44] F. Wang *et al.*, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3156–3164.
- [45] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
- [46] A. Pumarola, A. Agudo, A. M. Martinez, A. Sanfeliu, and F. Moreno-Noguer, "Ganimation: Anatomically-aware facial animation from a single image," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 818–833.
- [47] L. Chen *et al.*, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5659–5667.
- [48] X. Zhang, T. Wang, J. Qi, H. Lu, and G. Wang, "Progressive attention guided recurrent network for salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 714–722.
- [49] N. Liu, J. Han, and M.-H. Yang, "PiCANet: Learning pixel-wise contextual attention for saliency detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3089–3098.
- [50] S. Chen, X. Tan, B. Wang, and X. Hu, "Reverse attention for salient object detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 234–250.
- [51] W. Wang, S. Zhao, J. Shen, S. C. H. Hoi, and A. Borji, "Salient object detection with pyramid attention and salient edges," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1448–1457.
- [52] T. Zhao and X. Wu, "Pyramid feature attention network for saliency detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3085–3094.
- [53] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 694–711.
- [54] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.
- [55] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [56] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, 2017, pp. 214–223.
- [57] J. Zhao *et al.*, "Self-supervised neural aggregation networks for human parsing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 7–15.

- [58] Z. Jian, "Deep learning for human-centric image analysis," Ph.D. dissertation, Dept. Elect. Comput. Eng., National Univ. Singapore (NUS), Singapore, 2018.
- [59] J. Zhao, J. Li, Y. Cheng, T. Sim, S. Yan, and J. Feng, "Understanding humans in crowded scenes: Deep nested adversarial learning and a new benchmark for multi-human parsing," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 792–800.
- [60] J. Zhao, J. Li, H. Liu, S. Yan, and J. Feng, "Fine-grained multi-human parsing," *Int. J. Comput. Vis.*, vol. 128, nos. 8–9, pp. 2185–2203, Sep. 2020.
- [61] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2884–2896, Nov. 2018.
- [62] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 5767–5777.
- [63] K. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit. (FGR)*, 2006, pp. 341–345.
- [64] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset," *IEEE Trans. Multimedia*, vol. 17, no. 6, pp. 804–815, Jun. 2015.
- [65] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.
- [66] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 325–341.
- [67] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, "MaskGAN: Towards diverse and interactive facial image manipulation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5549–5558.



Yunfan Liu received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2015, and the M.S. degree in electronic engineering: systems from the University of Michigan at Ann Arbor, Ann Arbor, USA, in 2017. He is currently pursuing the Ph.D. degree with the National Laboratory of Pattern Recognition, Center for Research on Intelligent Perception and Computing, Institute of Automation, Chinese Academy of Sciences, Beijing. His research interests include computer vision, pattern recognition, and machine learning.



Qi Li (Member, IEEE) received the B.E. degree in automation from the China University of Petroleum, Qingdao, China, in 2011, and the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing, China, in 2016. He is currently an Associate Professor with the National Laboratory of Pattern Recognition, Center for Research on Intelligent Perception and Computing, CASIA. His research interests include face recognition, computer vision, and machine learning.



Zhenan Sun (Senior Member, IEEE) received the B.S. degree in industrial automation from the Dalian University of Technology, Dalian, China, in 1999, the M.S. degree in system engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2002, and the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing, China, in 2006. Since 2006, he has been a Faculty Member with the National Laboratory of Pattern Recognition, Center for Research on Intelligent Perception and Computing, CASIA, where he is currently a Professor. He has authored or coauthored more than 200 technical articles. His current research interests include biometrics, pattern recognition, and computer vision. He is a fellow of the IAPR. He serves as the Chair for IAPR Technical Committee on Biometrics. He is an Associate Editor of the IEEE TRANSACTIONS ON BIOMETRICS, BEHAVIOR, AND IDENTITY SCIENCE.



Tieniu Tan (Fellow, IEEE) received the B.S. degree in electronic engineering from Xi'an Jiaotong University, China, in 1984, and the M.S. and Ph.D. degrees in electronic engineering from Imperial College London, U.K., in 1986 and 1989, respectively. He is currently a Professor with the National Laboratory of Pattern Recognition, Center for Research on Intelligent Perception and Computing, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He has published more than 450 research papers in refereed international journals and conferences in the areas of image processing, computer vision, and pattern recognition, and has authored or edited 11 books. His research interests include biometrics, image and video understanding, information hiding, and information forensics. He is a fellow of the CAS, the TWAS, the BAS, the IAPR, the U.K. Royal Academy of Engineering, and the Past President of the IEEE Biometrics Council.