

# Attribute-aware Face Aging with Wavelet-based Generative Adversarial Networks

Yunfan Liu<sup>1\*</sup> Qi Li<sup>1,2,3\*</sup> Zhenan Sun<sup>1,2,4</sup>

<sup>1</sup> Center for Research on Intelligent Perception and Computing, CASIA

<sup>2</sup> National Laboratory of Pattern Recognition, CASIA

<sup>3</sup> Artificial Intelligence Research, CAS, Jiaozhou, Qingdao, China

<sup>4</sup> Center for Excellence in Brain Science and Intelligence Technology, CAS

yunfan.liu@cripac.ia.ac.cn, {qli, znsun}@nlpr.ia.ac.cn

## Abstract

Since it is difficult to collect face images of the same subject over a long range of age span, most existing face aging methods resort to unpaired datasets to learn age mappings. However, the matching ambiguity between young and aged face images inherent to unpaired training data may lead to unnatural changes of facial attributes during the aging process, which could not be solved by only enforcing identity consistency like most existing studies do. In this paper, we propose an attribute-aware face aging model with wavelet-based Generative Adversarial Networks (GANs) to address the above issues. To be specific, we embed facial attribute vectors into both the generator and discriminator of the model to encourage each synthesized elderly face image to be faithful to the attribute of its corresponding input. In addition, a wavelet packet transform (WPT) module is incorporated to improve the visual fidelity of generated images by capturing age-related texture details at multiple scales in the frequency space. Qualitative results demonstrate the ability of our model in synthesizing visually plausible face images, and extensive quantitative evaluation results show that the proposed method achieves state-of-the-art performance on existing datasets.

## 1. Introduction

Face aging, also known as age progression [16], aims at rendering a given face image with aging effects while still preserving personalized features. Applications of face aging techniques range from social security to digital entertainment, such as predicting contemporary appearance of missing children and cross-age identity verification. Due to the practical value of face aging, many approaches have

\*Authors contributed equally.

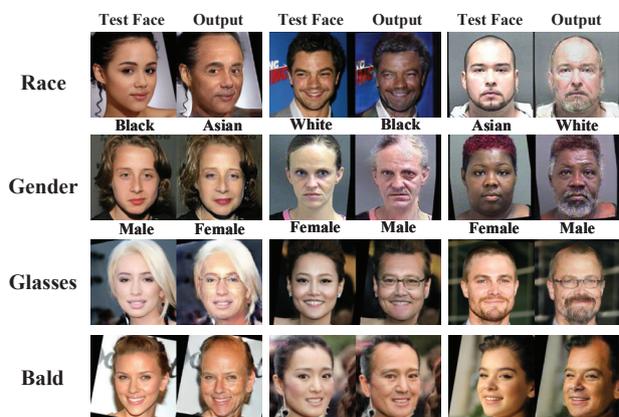


Figure 1. Examples of face aging with mismatched facial attributes generated by face aging model without facial attribute embedding. Four attributes (Race, Gender, Glasses, and Bald) are considered and three sample results are presented for each. Labels of ‘Race’ and ‘Gender’ are all obtained via advanced publicly available APIs of Face++ [13] and placed underneath each image.

been proposed to address this problem in the last two decades [8, 20, 19, 21, 7]. With the rapid development of deep learning, deep generative models are widely adopted to synthesize aged face images [23, 3, 4]. However, the most critical problem of these methods is that multiple face images of the same person at different ages are required at training stage, which is extremely expensive to collect in practice and thus their applications are largely limited.

To deal with this problem, many recent studies resort to unpaired face aging data to train the model [23, 28, 25, 9]. However, these approaches mainly focus on face aging itself while neglecting other critical conditional information of the input (*e.g.*, facial attributes), thus fail to regulate the training process. Consequently, training face image pairs with mismatched attributes would mislead the model

into learning translations other than aging, causing serious ghosting artifacts and even incorrect facial attributes in generation results. Fig. 1 shows some face aging results with mismatched attributes. In the rightmost face aging result under ‘gender’, beard is mistakenly attached to the input female face image. This is because the model learns that growing a beard is a typical sign of aging but has no way to know that this does not happen to a woman, as face image pairs of young woman and old man could be treated as positive training samples if no conditional information is considered.

To suppress such undesired changes of semantic information during the aging process, many recent face aging studies attempt to supervise the output by enforcing identity consistency [28, 1, 25, 9]. However, as shown in Fig. 1, personalized features are well preserved in the output for all sample results, nevertheless, obvious unnatural changes of facial attributes are still observed. In other words, well maintained identity-related features do NOT imply reasonable aging results when training with unpaired data. Therefore, merely enforcing identity consistency is insufficient to eliminate matching ambiguities in unpaired training data, thus fails to achieve satisfactory face aging performance.

To solve the above-mentioned issues, in this paper, we propose a framework based on generative adversarial networks (GANs). Different from existing methods in the literature, we involve semantic conditional information of the input by embedding facial attribute vectors in both the generator and discriminator, so that the model could be guided to output elderly face images with attributes faithful to each corresponding input. Furthermore, to enhance aging details, based on the observation that signs of aging are mainly represented by wrinkles, laugh lines, and eye bags, which could be treated as local textures, we employ wavelet packet transform to extract features at multiple scales in the frequency space efficiently.

To summarized, the main contributions are as follows:

- Facial attributes are incorporated as conditional information into both the generator and discriminator for face aging, since identity preservation is insufficient for generating reasonable results.
- Wavelet packet transform is adopted to extract features of texture details at multiple scales in the frequency domain for generating fine-grained details of aging effects.
- Extensive experiments have been conducted to demonstrate the ability of the proposed method in rendering accurate aging effects and preserving information of both identity and facial attributes. Quantitative results indicate that our method achieves state-of-the-art performance.

## 2. Related Work

In the last few decades, face aging has been a very popular research topic and a great amount of algorithms have been proposed to tackle this issue. In general, these methods could be divided into three categories: physical model-based methods, prototype-based methods, and deep learning-based methods.

Physical model-based methods mechanically simulate the changes of facial appearance w.r.t. time by modeling the anatomical structure of human faces. Todd et al. [22] modeled the translation of facial appearance by revised cardioid strain transformation. Subsequent works investigated the problem from various biological aspects including muscles and overall facial structures [8, 20]. However, physical model-based algorithms are computational expensive and large amount of image sequences of the same subject are required to model aging effects.

Data-driven prototyping approaches [19, 21, 7] come into view the next, where faces are divided into age groups and each group is represented by an average face (prototype) computed from the training data. After that, translation patterns between prototypes are regarded as effects of aging. The main problem of prototyping methods is that personalized features are eliminated when calculating average faces, thus the identity information is not well preserved.

In recent years, deep generative models with temporal architectures are adopted to synthesize images of elderly faces [23, 3, 4]. However, in most of these works, face image sequence over a long age span for each subject is required thus their potential in practical use is limited. With the success of GANs [5] in generating visually appealing images, many efforts have been made to tackle the problem of face aging using GAN-based framework [28, 25, 9, 17, 24, 10]. Zhang *et al.* [28] proposed a conditional adversarial autoencoder (CAAE) to achieve age progression and regression by traversing in low-dimensional manifold. The work most similar to ours is [25], in which a GAN-based model with pyramid architecture is proposed, and identity loss is adopted to achieve permanence. Besides preserving identity information, we focus on alleviating the influence of matching ambiguity of unpaired training samples and ensuring attribute consistency by embedding facial attribute vectors in the model.

## 3. Approach

In a unpaired face aging dataset, each young face image might map to many elderly face candidates during the training process, and image pairs with mismatched semantic information may mislead the model into learning translations other than aging. To solve this problem, we present a GAN-based face aging model that takes young face images

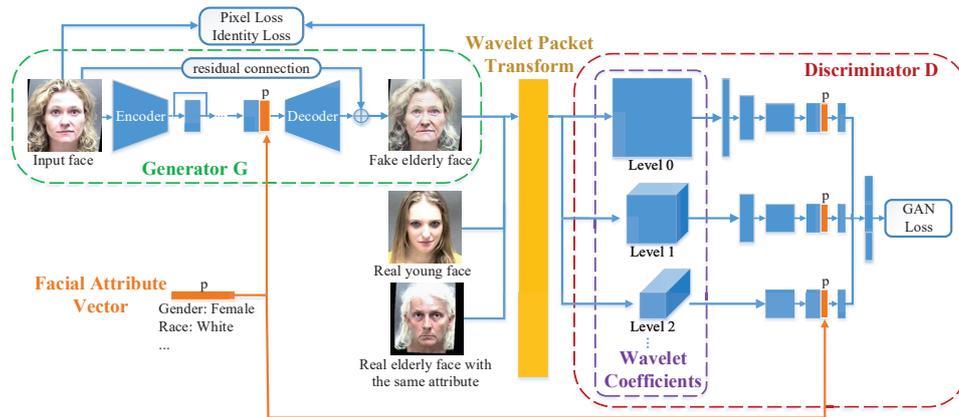


Figure 2. An overview of the proposed face aging framework. An hourglass-shaped generator  $G$  learns the age mapping and outputs lifelike elderly face images. A discriminator  $D$  is employed to distinguish synthesized face images from generic ones, based on multi-scale wavelet coefficients computed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

and their semantic information (*i.e.* facial attributes) as input and outputs visually plausible aged faces accordingly. The network consists of two parts: a facial attribute embedded generator  $G$  and a wavelet-based discriminator  $D$ . The generator network embeds facial attributes into young face images and synthesizes aged faces. The discriminator network is used to encourage the generation results to be indistinguishable from generic ones and to possess attributes same as the corresponding input. An overview of the proposed framework is presented in Fig. 2.

### 3.1. Facial Attribute Embedded Generator

Existing face aging studies [9, 25, 28] only take young face images as inputs and then directly learn age mappings using GAN-based networks. Although constraints on identity information and pixel values are usually imposed to restrict modifications made to input images, facial attributes may still undergo unnatural translations (as shown in Fig. 1). Unlike previous works, we propose to incorporate both low-level image information (pixel values) and high-level semantic information (facial attributes) into the face aging model to regularize image translation patterns and reduce the ambiguity of mappings between unpaired young and aged faces. To be specific, the model takes young face images and their corresponding attribute vectors as input, and generates elderly face images with attributes in agreement with the input ones.

Rather than supervising the attributes of generation results by simply adopting an additional loss term, we embed the attribute vector in the generator so that semantic facial information is well considered in the generation process and encourages the model to produce face images with consistent attributes more effectively. To be specific, we employ

an hourglass-shaped fully convolutional network as the generator, which has achieved success in previous image translation studies [6, 29]. It consists of an encoder network, a decoder network, and four residual blocks in between as the bottleneck. The input facial attribute vector is replicated and concatenated to the output blob of the last residual block as they both contain high-level semantic features. After the combination, the decoder network transforms the concatenated feature blob back to the image space.

Since face aging could be considered as rendering aging effects conditioned on the input young face image, we add the input image to the output of the decoder to form a residual connection. Compared to synthesize the whole face image, this structure automatically makes the generator focus more on modeling the difference between input and output face images, namely the representative signs of aging, and be less likely to be distracted by visual content irrelevant to aging, such as background. Finally, the numeric scale of the resultant tensor is normalized by a hyperbolic tangent ( $\tanh$ ) mapping and thus the generated elderly face image is obtained.

### 3.2. Wavelet-based Discriminator

To force the generator to absorb the semantic information of the input face image, a conditional discriminator is employed. The discriminator has two main functions: 1) distinguish synthesized face images from generic ones; 2) check whether the attribute of each generation result is faithful to that of the corresponding input.

To be specific, considering the fact that typical signs of aging, such as wrinkles, laugh lines, and eye bags, could be regarded as local image textures, we adopt wavelet packet transform (WPT, see Fig. 3) to capture age-related textu-

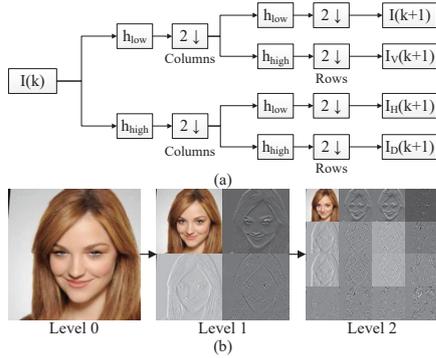


Figure 3. Demonstration of wavelet packet transform. (a) Low-pass and high-pass decomposition filters ( $h_{low}$  and  $h_{high}$ ) are applied iteratively to the input on  $k$ -th level to compute wavelet coefficients on the next level; (b) a sample face image with its wavelet coefficients at different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

ral features. Specifically, multi-level WPT is performed to provide a more comprehensive analysis of textures in the given image, and wavelet coefficients at each decomposing level are fed into a convolutional pathway of the discriminator. Note that this is different from [9], since wavelet coefficients are only used for discrimination in our work and no prediction or reconstruction is involved.

To make the discriminator gain the ability of telling whether attributes are preserved in generated images, the input attribute vector is also replicated and concatenated to the output of an intermediate convolutional block of each pathway. At the end of the discriminator, same-sized outputs of all pathways are fused to form a single tensor, and adversarial loss is then estimated against the label tensor.

Compared to extracting multi-scale features by a sequence of convolutional layers as in [25], the advantage of using WPT is that the computational cost is significantly reduced since calculating wavelet coefficients could be regarded as 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, it still takes the advantage of multi-scale image texture analysis, which is helpful in improving the visual fidelity of generated images.

### 3.3. Overall Objective Functions

Training of GAN model simulates the process of optimizing a minimax-max two-player game between the generator  $G$  and the discriminator  $D$ . Unlike regular GANs [5], we adopt least square loss instead of negative log likelihood loss for that margins between generated samples and the decision boundary in the feature space are also minimized, which further improves the quality of synthesized images [12]. Practically, we pair up young face images  $x_i$  and their corresponding attribute vectors  $\alpha_i$  of dimen-

sion  $p$ , denoted as  $(x_i, \alpha_i) \sim P_{young}(x, \alpha)$ , and take them as input to the model. Only generic aged faces with attributes same as the input, i.e.  $(x_i, \alpha_i) \sim P_{old}(x, \alpha_i)$ , are considered as positive samples, and real young faces, i.e.  $(x_i, \alpha_i) \sim P_{young}(x, \alpha)$ , are regarded as negative samples to help  $D$  gain discriminating ability on aging effects.

Mathematically, the objective function for  $G$  and  $D$  could be written as follows,

$$L_{GAN}(G) = \mathbb{E}_{(x_i, \alpha_i) \sim P_{young}(x, \alpha)} [(D(G(x_i, \alpha_i), \alpha_i) - 1)^2] \quad (1)$$

$$L_{GAN}(D) = \mathbb{E}_{(x_i, \alpha_i) \sim P_{old}(x, \alpha_i)} [(D(x_i, \alpha_i) - 1)^2] + \mathbb{E}_{(x_i, \alpha_i) \sim P_{young}(x, \alpha)} [D(G(x_i, \alpha_i), \alpha_i)]^2 + \mathbb{E}_{(x_i, \alpha_i) \sim P_{young}(x, \alpha)} [D(x_i, \alpha_i)]^2 \quad (2)$$

where  $P_{young}$  and  $P_{old}$  denote the distribution of generic face images of young and old subjects, respectively.

In addition, pixel loss and identity loss are adopted to maintain consistency in both image-level and personalized feature-level. To be specific, we utilize the VGG-Face descriptor [14], denoted by  $\phi$ , to extract the identity related semantic representation of a face image. These two loss terms could be formulated as,

$$L_{pix} = \mathbb{E}_{(x_i, \alpha_i) \sim P_{young}(x, \alpha)} \|G(x_i, \alpha_i) - x_i\|_F^2 \quad (3)$$

$$L_{id} = \mathbb{E}_{(x_i, \alpha_i) \sim P_{young}(x, \alpha)} \|\phi(G(x_i, \alpha_i)) - \phi(x_i)\|_F^2 \quad (4)$$

In conclusion, overall objective functions of the proposed model could be written as follows,

$$L_G = L_{GAN}(G) + \lambda_{pix} L_{pix} + \lambda_{id} L_{id} \quad (5)$$

$$L_D = L_{GAN}(D) \quad (6)$$

where  $\lambda_{id}$  and  $\lambda_{pix}$  are coefficients balancing the importance of critics on identity and pixels, respectively. We optimize the model by minimizing  $L_G$  and  $L_D$  alternatively until the optimality is reached.

## 4. Experiments

### 4.1. Dataset

MORPH [15] is a large aging dataset containing 55,000 face images of more than 13,000 subjects. Data samples in MORPH are color images of near-frontal faces exhibiting neutral expressions under uniform and moderate illumination with simple background. CACD [2] contains 163,446 face images of 2,000 celebrities captured in much less controlled conditions. Besides large variations in pose, illumination, and expression (PIE variations), images in CACD are collected via Google Image Search, making it a very challenging dataset due to the mismatching between actual face presented in each image and associated labels provided (name and age).

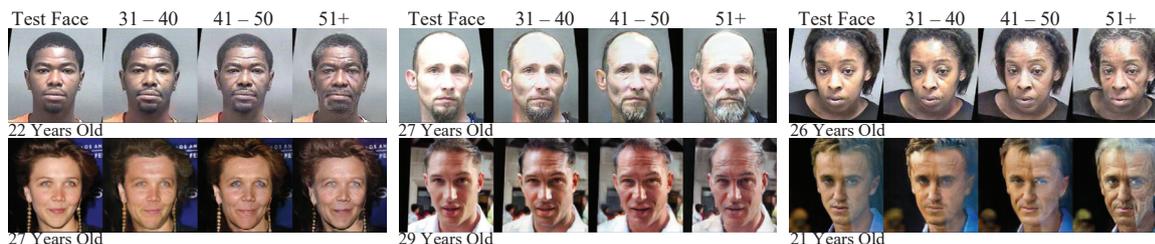


Figure 4. Sample results on Morph (first row) and CACD (second row). The first image in each result is the input test face image and subsequent 3 images are synthesized elderly face images of the same subject in age group 31-40, 41-50 and 51+, respectively.

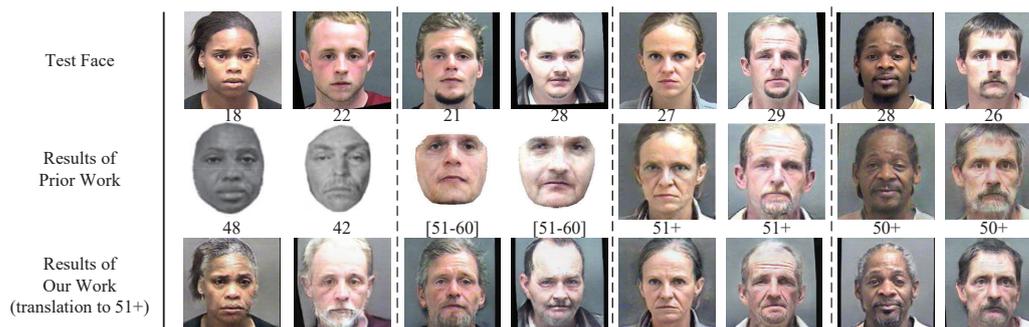


Figure 5. Performance comparison with prior work on Morph (zoom in for a better view of the aging details). The second row shows the results of prior work, where four methods are considered and two sample results are presented for each. These four methods are (from left to right): CONGRE [18], HFA [26], GLCA-GAN [9], and PAG-GAN [25]. The last row shows the results of our method.

As for facial attributes, MORPH provides researchers with labels including age, gender, and race for each image. We choose ‘gender’ and ‘race’ to be the attributes that are required to be preserved, since these two attributes are guaranteed to remain unchanged during natural aging process, and are relatively objective compared to attributes such as ‘attractive’ or ‘chubby’ used in popular facial attribute dataset CelebA [11]. For CACD, since face images with race other than ‘white’ only takes a small portion of the entire dataset, we only select ‘gender’ as the attribute to preserve. To be specific, we go through the name list of the celebrities and label the corresponding images accordingly. This introduces noise in gender labels due to the mismatching between the annotated name and the actual face presented in each image, which further increases the difficulty for our method to achieve good performance on this dataset. It is worthwhile to note that the proposed model is highly expandable, as researchers may choose whatever attributes to preserve simply by incorporating them in the conditional facial attribute vector and arrange training images pairs accordingly.

## 4.2. Implementation Details

All face images are cropped and aligned according to the five facial landmarks detected by MTCNN [27]. Following the convention in [25, 9], we divide the face images into four age groups, *i.e.*, 30-, 31-40, 41-50, 51+, and only con-

sider translations from 30- to the other three age groups. To evaluate the performance of the proposed method objectively, all metric measurements are conducted via stable public APIs of Face++ [13]. Thresholds adopted in our face verification experiments (threshold=76.5, FAR=1e-5) are the same as those used in [25]. Therefore, quantitative results of our experiments are comparable to those reported in [25].

We choose Adam to be the optimizer of both  $G$  and  $D$  with learning rate and batch-size set to  $1e^{-4}$  and 16, respectively. Pixel-level critic is applied every 5 iterations, and  $D$  is updated at every iteration. As for trade-off parameters,  $\lambda_{pix}$  and  $\lambda_{id}$  are firstly set to make  $L_{pix}$  and  $L_{id}$  to be of the same order of magnitude as  $L_{GAN}(G)$ , and then divided by 10 to emphasize the importance of the adversarial loss. All experiments are conducted under 5-fold cross validation on a Nvidia Titan Xp GPU.

## 4.3. Qualitative Results of Face Aging

Sample results on Morph and CACD are shown in Fig. 4. It is clear that our method is able to simulate translations between age groups and synthesize elderly face images with high visual fidelity. In addition, our method is robust to variations in terms of race, gender, expression, and occlusion.

Performance comparison with prior work on Morph is shown in Fig. 5. Traditional face aging methods, CONGRE [18] and HFA [26], only render subtle aging effects

Table 1. Age estimation results on Morph and CACD (differences of mean ages are measured in absolute value).

Morph				CACD			
Age group	31 - 40	41 - 50	51 +	Age group	31 - 40	41 - 50	51 +
Estimated Age Distributions				Estimated Age Distributions			
Generic	38.60	47.74	57.25	Generic	38.51	46.54	53.39
Synthetic	38.47	47.55	56.57	Synthetic	38.88	47.42	54.05
Difference of mean ages				Difference of mean ages			
CAAE	10.08	15.49	21.42	CAAE	5.76	11.53	17.93
GLCA-GAN	0.23	3.61	8.61	GLCA-GAN	1.72	2.07	2.85
PAG-GAN	0.38	0.52	1.48	PAG-GAN	0.70	<b>0.22</b>	<b>0.57</b>
Ours	<b>0.13</b>	<b>0.19</b>	<b>0.68</b>	Ours	<b>0.37</b>	0.58	0.66

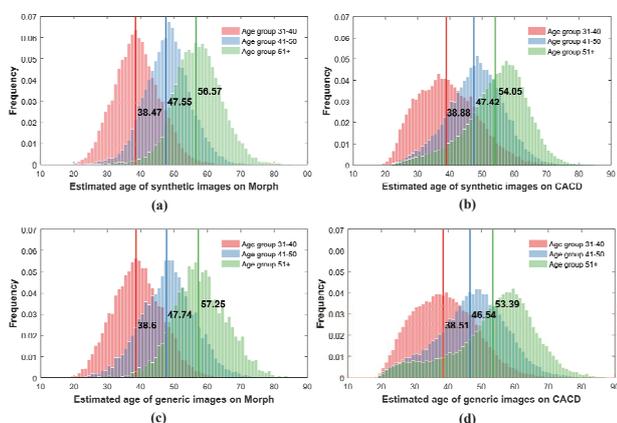


Figure 6. Distributions of the estimated ages. (a) synthetic faces on Morph; (b) synthetic faces on CACD; (c) generic faces on Morph; (d) generic faces on CACD.

within tight facial area, which fails to accurately simulate the aging process. In contrast, GAN-based methods, GLCA-GAN [9] and GAN with pyramid architecture proposed in [25], referred to as PAG-GAN, have achieved significant improvement on the quality of generation results. However, our method further generates face images of higher resolution ( $2\times$ ) with enhanced details compared to GLCA-GAN, and reduces ghosting artifacts in the results compared to PAG-GAN (*e.g.* finer details of hair and beard).

#### 4.4. Aging Accuracy and Identity Preservation

In this subsection, we report evaluation results on aging accuracy and identity preservation. The performance of the proposed model is compared with previously state-of-the-art methods CAAE [28], GLCA-GAN [9] and PAG-GAN [25] to demonstrate the effectiveness.

**Aging Accuracy:** Age distributions of both generic and synthetic faces in each age group are estimated, where less discrepancy between real and fake images indicates more accurate simulation of aging effects. On Morph and CACD,

face images of age under or equal to 30 are considered as testing samples, and their corresponding aged faces in the other three age groups are synthesized. We estimated the apparent age of both generation results and natural face images in the dataset using Face ++ APIs for fair comparison.

Age estimation results on Morph and CACD are shown in Table 1 and Fig. 6. We compare our method with previous works in terms of differences between mean ages. On Morph, it could be seen that estimated age distributions of synthetic elderly face images well match that of natural images for all age groups. Our method consistently outperforms other approached in all three aging processes, demonstrating the effectiveness of our method. Signs of aging in results of CAAE are not obvious enough, leading to large age estimation errors. On CACD, due to the existence of mismatching between face images and associated labels, slight performance drop could be observed. Still, the proposed method achieves results comparable to previous state-of-the-art. This shows that our method is relatively robust to noise in attribute labels and thus lower the requirement on the accuracy of the prior attribute detection process.

**Identity Preservation:** Face verification experiments are conducted to check whether the identity information has been preserved during the face aging process. Similar to previous literature, comparisons between synthetic elderly face images from different age groups of the same subject are also conducted to inspect if the identity information is consistent among three separately trained age mappings.

Results of face verification experiments are shown in Table 2. On Morph, our method achieves the highest verification rate on all three translations and outperforms other approaches by a clear margin, especially in the hardest case (from 30- to 51+). This demonstrates that the proposed method successfully achieves identity permanence during face aging. On the more challenging dataset CACD containing mismatched labels, the performance of our method is comparable to PAG-GAN with minor difference. Notably, as the time interval between two face images of a single subject increases, both verification confidence and accuracy

Table 2. Face verification results on Morph and CACD.

Morph				CACD			
Age group	31 - 40	41 - 50	51 +	Age group	31 - 40	41 - 50	51 +
Verification Confidence				Verification Confidence			
30 -	95.77	94.64	87.53	30 -	93.67	91.54	90.32
31 - 40	-	95.47	89.53	31 - 40	-	91.74	90.54
41 - 50	-	-	90.50	41 - 50	-	-	91.12
Verification Rate (%)				Verification Rate (%)			
CAAE	15.07	12.02	8.22	CAAE	4.66	3.41	2.40
GLCA-GAN	97.66	96.67	91.85	GLCA-GAN	97.72	94.18	92.29
PAG-GAN	<b>100.00</b>	98.91	93.09	PAG-GAN	<b>99.99</b>	<b>99.81</b>	98.28
Ours	<b>100.00</b>	<b>100.00</b>	<b>98.26</b>	Ours	99.76	98.74	<b>98.44</b>

Table 3. Facial attributes preservation rates for ‘Gender’ and ‘Race’ on Morph and CACD.

Age group	Preservation Rate (%) of ‘Gender’						Preservation Rate (%) of ‘Race’		
	Morph			CACD			Morph		
	31 - 40	41 - 50	51 +	31 - 40	41 - 50	51 +	31 - 40	41 - 50	51 +
GLCA-GAN	96.30	95.43	95.77	87.27	86.79	85.89	91.79	89.52	89.34
PAG-GAN	95.96	93.77	92.47	83.97	81.28	70.05	95.83	88.51	87.98
Ours	<b>97.37</b>	<b>97.21</b>	<b>96.07</b>	<b>90.71</b>	<b>87.63</b>	<b>87.19</b>	<b>95.86</b>	<b>94.10</b>	<b>93.22</b>

decrease, which is reasonable as greater changes in facial appearance may occur as more time elapsed.

#### 4.5. Facial Attribute Consistency

We evaluate the performance of facial attribute preservation by comparing facial attributes estimated before and after age progression, and results are listed in Table 3. On Morph, facial attributes of the majority of testing samples (up to 97.37% for ‘gender’ and 95.86% for ‘race’) are well preserved in the aging process. In addition, our method outperforms both GLCA-GAN and PAG-GAN by clear margins on translations to all age groups. On CACD, due to the influence of mistakenly labeled data samples, clear performance drop could be observed compared to the results on Morph. However, our method still gives better performance on facial attributes preservation than other methods. The advantage of our method in preserving the ‘gender’ attribute becomes greater as the age gap increases, and finally reaches 17.14% (87.19% over 70.05%) when translating to the oldest age group 51+. From Table 3, we could conclude that undesired changes of facial attributes are more likely to happen as the age gap increases, and incorporating conditional information is beneficial for maintaining consistency of target facial attributes in the aging process.

#### 4.6. Ablation Study

In this part, experiments are conducted to fully explore the contribution of facial attribute embedding (FAE)



Figure 7. Sample visual results of the ablation study. For each face, the estimated age (first row) and detected attributes (second row) are listed underneath. Values in the last row are face verification confidence between generation results and the test face.

and wavelet packet transform (WPT) in simulating accurate age translations. We investigate the impact of including/excluding attribute embedding (w/wo FAE) and wavelet packet transform (w/wo WPT) on age distribution, face verification rate, and attribute preservation rate. All experiments in this subsection are conducted only on Morph as labels are noisy on CACD dataset.

Visual illustrations of face images generated by variants of the proposed model are shown in Fig. 7. It is clear that when both FAE and WPT are not involved (woFAE\_woWPT), generation results suffer from severe

Table 4. Comparison of results on facial attribute preservation and aging accuracy between variants of the proposed model (differences of mean ages are measured in absolute value).

Age group	Gender Preservation Rate (%)			Race Preservation Rate (%)			Deviation of Estimated Ages		
	31-40	41-50	51+	31-40	41-50	51+	31-40	41-50	51+
woFAE / woWPT	95.72	94.21	93.60	95.04	93.55	90.83	0.44	1.72	3.03
woFAE / wWPT	96.15	94.90	93.61	93.89	88.63	90.21	0.68	0.41	2.31
wFAE / woWPT	97.21	96.91	95.85	95.22	<b>94.35</b>	91.43	0.82	0.52	4.82
Ours	<b>97.37</b>	<b>97.21</b>	<b>96.07</b>	<b>95.86</b>	94.10	<b>93.22</b>	<b>0.13</b>	<b>0.19</b>	<b>0.68</b>

Table 5. Face verification rates (%) of variants of the proposed model on Morph

Age group	31-40	41-50	51+
woFAE / woWPT	100.00	100.00	99.92
woFAE / wWPT	100.00	99.88	98.06
wFAE / woWPT	100.00	100.00	98.86
Ours	100.00	100.00	98.26

ghosting artifacts. Due to the intrinsic matching ambiguity of unpaired training data, the model without FAE mistakenly attaches moustache to the input female face image to show the aging effect. Notably, growing a moustache does not decrease the face verification confidence, as the generated face image still shares similar identity-related features with the input. This again confirms our observation that enforcing identity consistency is insufficient to obtain satisfactory face aging results.

On the contrary, incorporating FAE suppresses the undesired facial attribute drift by reducing the matching ambiguity. To be specific, in Fig. 7, there is no more moustache in generation results after adopting FAE thus facial attribute consistency is achieved. Unfortunately, removing moustache also wipes out aging-related textural details (wrinkles, laugh lines, and eye bags), leading to relatively inaccurate aging results (much younger than expected).

To solve this issue and generate more visually plausible face images with vivid signs of aging, WPT is employed as the initial layer of the discriminator. The contribution of WPT could be easily seen by comparing the results obtained under setting ‘woFAE / woWPT’ and ‘woFAE / wWPT’, as well as ‘wFAE / woWPT’ and ‘Ours’. Although results obtained under setting ‘woFAE / wWPT’ still suffer from wrong facial attributes, ghosting artifacts are significantly alleviated and lifelike aging effects are clearly observed.

Quantitative results for ablation study are shown in Table 4 and 5. According to results in Table 4, introducing facial attribute embedding (wFAE) increases preservation rates for both ‘gender’ and ‘race’ under all three age mappings, especially in the case of translating to 51+. This proves the effectiveness of attribute embedding as it aligns unpaired age data in terms of facial attributes and thus re-

duces the intrinsic ambiguity in data mapping.

In addition, it is clear that adopting WPT reduces the discrepancies between age distributions of generic and synthetic images in all cases. However, WPT provides little help in maintaining facial attribute consistency. This is because WPT only captures feature based on low-level visual data and could not bridge the semantic gap, so that the framework still suffers from mismatched data samples.

Combining results in Table 4 and 5, it could be seen that while attribute preservation rates still have room for improvement, verification rates are about to reach perfection. This observation validates our statement that identity preservation does not guarantee that facial attributes remain stable during the aging process. Therefore, besides constraints on identity, supervision on facial attributes are also helpful to reduce the intrinsic matching ambiguity of unpaired data and achieve satisfactory face aging results.

## 5. Conclusion

In this paper, we propose a GAN-based framework to synthesize aged face images. Due to the ineffectiveness of identity constraints in reducing the matching ambiguity of unpaired aging data, we propose to employ facial attributes to tackle this issue. Specifically, we embed facial attribute vectors to both the generator and discriminator to encourage generated images to be faithful to facial attributes of the corresponding input image. To further improve the visual fidelity of generated face images, wavelet packet transform is introduced to extract textual features at multiple scales efficiently. Extensive experiments are conducted on Morph and CACD, and qualitative results demonstrate that our method could synthesize lifelike face images robust to both PIE variations and noisy labels. Furthermore, quantitative results obtained via public APIs validate the effectiveness of the proposed method in aging accuracy as well as identity and attribute preservation.

**Acknowledgements.** This work is supported by the National Natural Science Foundation of China (Grant No. 61702513, U1836217, 61427811).

## References

- [1] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. Face aging with conditional generative adversarial networks. *IEEE International Conference on Image Processing (ICIP)*, pages 2089–2093, 2017.
- [2] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Transactions on Multimedia (TMM)*, 17(6):804–815, 2015.
- [3] Chi Nhan Duong, Khoa Luu, Kha Gia Quach, and Tien D Bui. Longitudinal face modeling via temporal deep restricted boltzmann machines. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5772–5780, 2016.
- [4] Chi Nhan Duong, Kha Gia Quach, Khoa Luu, T Hoang Ngan Le, and Marios Savvides. Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 3755–3763, 2017.
- [5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems (NIPS)*, pages 2672–2680, 2014.
- [6] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision (ECCV)*, pages 694–711, 2016.
- [7] Ira Kemelmacher-Shlizerman, Supasorn Suwajanakorn, and Steven M Seitz. Illumination-aware age progression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3334–3341, 2014.
- [8] Andreas Lanitis, Christopher J. Taylor, and Timothy F Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 24(4):442–455, 2002.
- [9] Peipei Li, Yibo Hu, Qi Li, Ran He, and Zhenan Sun. Global and local consistent age generative adversarial networks. In *International Conference on Pattern Recognition (ICPR)*, pages 1073–1078, 2018.
- [10] Qi Li, Yunfan Liu, and Zhenan Sun. Age progression and regression with spatial attention modules. In *arXiv preprint arXiv:1903.02133*, 2019.
- [11] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.
- [12] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2813–2821, 2017.
- [13] Megvii Inc. Face++ research toolkit. <http://www.faceplusplus.com>.
- [14] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. Deep face recognition. In *British Machine Vision Conference (BMVC)*, pages 41.1–41.12, 2015.
- [15] Karl Ricanek and Tamirat Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *the International Conference on Automatic Face and Gesture Recognition (FG)*, pages 341–345, 2006.
- [16] Xiangbo Shu, Jinhui Tang, Hanjiang Lai, Luoqi Liu, and Shuicheng Yan. Personalized age progression with aging dictionary. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 3970–3978, 2015.
- [17] Jingkuan Song, Jingqiu Zhang, Lianli Gao, Xianglong Liu, and Heng Tao Shen. Dual conditional gans for face aging and rejuvenation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*, pages 899–905, 2018.
- [18] Jinli Suo, Xilin Chen, Shiguang Shan, Wen Gao, and Qionghai Dai. A concatenational graph evolution aging model. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 34(11):2083–2096, 2012.
- [19] Jinli Suo, Song-Chun Zhu, Shiguang Shan, and Xilin Chen. A compositional and dynamic model for face aging. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32(3):385–401, 2010.
- [20] Yusuke Tazoe, Hiroaki Gohara, Akinobu Maejima, and Shigeo Morishima. Facial aging simulator considering geometry and patch-tiled texture. In *ACM SIGGRAPH*, 2012.
- [21] Bernard Tiddeman, Michael Burt, and David Perrett. Prototyping and transforming facial textures for perception research. *IEEE Computer graphics and applications*, 21(5):42–50, 2001.
- [22] James T Todd, Leonard S Mark, Robert E Shaw, and John B Pittenger. The perception of human growth. *Scientific American*, 242(2):132–145, 1980.
- [23] Wei Wang, Zhen Cui, Yan Yan, Jiashi Feng, Shuicheng Yan, Xiangbo Shu, and Nicu Sebe. Recurrent face aging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2378–2386, 2016.
- [24] Z. Wang, W. Luo X. Tang, and S. Gao. Face aging with identity-preserved conditional generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [25] Hongyu Yang, Di Huang, Yunhong Wang, and Anil K. Jain. Learning face age progression: A pyramid architecture of gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 31–39, 2018.
- [26] Hongyu Yang, Di Huang, Yunhong Wang, Heng Wang, and Yuanyan Tang. Face aging effect simulation using hidden factor analysis joint sparse representation. *IEEE Transactions on Image Processing (TIP)*, 25(6):2493–2507, 2016.
- [27] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.
- [28] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4352–4360, 2017.

- [29] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017.