Learning Face Age Progression: A Pyramid Architecture of GANs

Hongyu Yang[1] Di Huang[1] Yunhong Wang[1] Anil K. Jain[2]
[1]Beihang University, China
[2]Michigan State University, USA
{hongyuyang, dhuang, yhwang}@buaa.edu.cn, jain@cse.msu.edu

Abstract

The two underlying requirements of face age progression, i.e. aging accuracy and identity permanence, are not well handled in the literature. In this paper, we present a novel generative adversarial network based approach. It separately models the constraints for the intrinsic subject-specific characteristics and the age-specific facial changes with respect to the elapsed time, ensuring that the generated faces present desired aging effects while simultaneously keeping personalized properties stable. Further, to generate more lifelike facial details, high-level age-specific features conveyed by the synthesized face are estimated by a pyramidal adversarial discriminator at multiple scales, which simulates the aging effects in a finer manner. The proposed method is applicable for diverse face samples in the presence of variations in pose, expression, makeup, etc., and remarkably vivid aging effects are achieved. Both visual fidelity and quantitative evaluations show that the approach advances the state-of-the-art.

1. Introduction

Age progression is the process of aesthetically rendering a given face image to present the effects of aging. It is often used for entertainment and forensics, *e.g.*, forecasting facial appearances of young children when they grow up or generating contemporary photos for missing individuals.

The intrinsic complexity of physical aging, the interferences caused by other factors (*e.g.*, PIE variations), and shortage of labeled aging data collectively make face age progression a rather difficult problem. The last few years have witnessed significant efforts tackling this issue, where aging accuracy and identity permanence are commonly regarded as the two underlying premises of its success [28][35][25][13]. The early attempts were mainly based on the skin's anatomical structure and they mechanically simulated the profile growth and facial muscle changes w.r.t. the elapsed time [30][34][22]. These methods provided the first insight into face aging synthesis. However, they generally worked in a complex manner, making it difficult to generalize. Data-driven approaches followed, where face age progression was primarily carried out by applying the prototype of aging details to test faces [12][28], or by modeling the dependency between longitudinal facial changes and corresponding ages [27][33][19]. Although obvious signs of aging are synthesized, their aging functions usually cannot formulate the complex aging mechanism accurately enough, limiting the diversity of aging patterns.

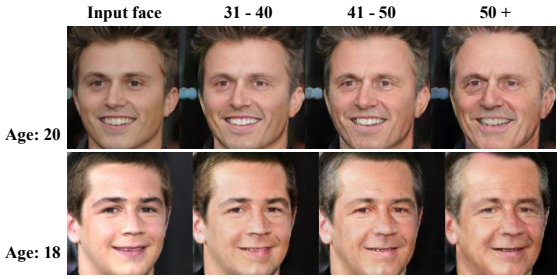

Figure 1. Demonstration of our aging simulation results (images in the first column are input faces of two subjects).

The deep generative networks have exhibited a remarkable capability in image generation [7][8][10][29] and have also been investigated for age progression [32][36][17][18]. These approaches render faces with more appealing aging effects and less ghosting artifacts compared to the previous conventional solutions. However, the problem has not been essentially solved. Specifically, these approaches focus more on modeling face transformation between two age groups, where the age factor plays a dominant role while the identity information plays a subordinate role, with the result that aging accuracy and identity permanence can hardly be simultaneously achieved, in particular for long-term age progression [17][18]. Furthermore, they mostly require multiple face images of different ages of the same individual at the training stage, involving another intractable issue, *i.e.* intra-individual aging face sequence collection [32][14]. Both the aforementioned facts indicate that current deep generative aging methods leave space for improvement.

In this study, we propose a novel approach to face age

progression, which integrates the advantage of Generative Adversarial Networks (GAN) in synthesizing visually plausible images with prior domain knowledge in human aging. Compared with existing methods in literature, it is more capable at handling the two critical requirements in age progression, *i.e.* identity permanence and aging accuracy. To be specific, the proposed approach uses a Convolutional Neural Networks (CNN) based generator to learn age transformation, and it separately models different face attributes depending upon their change over time. The training critic thus incorporates the squared Euclidean loss in the image space, the GAN loss that encourages generated faces to be indistinguishable from the elderly faces in the training set in terms of age, and the identity loss which minimizes the input-output distance by a high-level feature representation embedding personalized characteristics. It ensures that the resulting faces present desired effects of aging while the identity properties remain stable. By estimating the data density of each individual target age cluster, our method does not demand matching face pairs of the same person across two age domains as the majority of the counterpart methods do. Additionally, in contrast to the previous techniques that primarily operate on cropped facial areas (usually excluding foreheads), we emphasize that synthesis on entire faces is important since the parts of forehead and hair also significantly impact the perceived age. To achieve this and further enhance the aging details, our method leverages the intrinsic hierarchy of deep networks, and a discriminator of the pyramid architecture is designed to estimate high-level age-related clues in a fine-grained way. Our approach overcomes the limitations of single age-specific representation and handles age transformation both locally and globally. As a result, more photorealistic imageries are generated (see Fig. 1 for an illustration of aging results).

The main contributions of this study include:

1 We propose a novel GAN based method for age progression, which incorporates face verification and age estimation techniques, thereby addressing the issues of aging effect generation and identity cue preservation in a coupled manner.
2 We highlight the importance of the forehead and hair components of a face that are closely related to the perceived age but ignored in other studies; it indeed enhances the synthesized age accuracy.
3 We set up new validating experiments in addition to existent ones, including commercial face analysis tool based evaluation and insensitivity assessment to the changes in expression, pose, and makeup. Our method is not only shown to be effective but also robust in age progression.

2. Related Work

In the initial explorations of face age progression, physical models were exploited to simulate the aging mechanisms of cranium and facial muscles. In [30], Todd *et al.* introduced a revised cardioidal-strain transformation where head growth was modeled in a computable geometric procedure. Based on skin's anatomical structure, Wu *et al.* [34] proposed a 3-layered dynamic skin model to simulate wrinkles. Mechanical aging methods were also incorporated by Ramanathan and Chellappa [22] and Suo *et al.* [27].

The majority of the subsequent approaches were data-driven, which did not rely much on the biological prior knowledge, and the aging patterns were learned from the training faces. Wang *et al.* [33] built the mapping between corresponding down-sampled and high-resolution faces in a tensor space, and aging details were added on the later. Kemelmacher-Shlizerman *et al.* [12] presented a prototype based method, and further took the illumination factor into account. Yang *et al.* [35] first settled the multi-attribute decomposition problem, and progression was achieved by transforming only the age component to a target age group. These methods did improve the results, however ghosting artifacts frequently appeared on the synthesized faces.

More recently, the deep generative networks have been attempted. In [32], Wang *et al.* transformed faces across different ages smoothly by modeling the intermediate transition states in an RNN model. But multiple face images of various ages of each subject were required at the training stage, and the exact age label of the probe face was needed during test, thus greatly limiting its flexibility. Under the framework of conditional adversarial autoencoder [36], facial muscle sagging caused by aging was simulated, whereas only rough wrinkles were rendered mainly due to the insufficient representation ability of the training discriminator. With the Temporal Non-Volume Preserving (TNVP) aging approach [17], the short-term age progression was accomplished by mapping the data densities of two consecutive age groups with ResNet blocks [9], and the long-term aging synthesis was finally reached by a chaining of short-term stages. Its major weakness, however, was that it merely considered the probability distribution of a set of faces without any individuality information. As a result, the synthesized faces in a complete aging sequence varied a lot in color, expression, and even identity.

Our study also makes use of the image generation ability of GAN, and presents a different but effective method, where the age-related GAN loss is adopted for age transformation, the individual-dependent critic is used to keep the identity cue stable, and a multi-pathway discriminator is applied to refine aging detail generation. This solution is more powerful in dealing with the core issues of age progression, *i.e.* age accuracy and identity preservation.

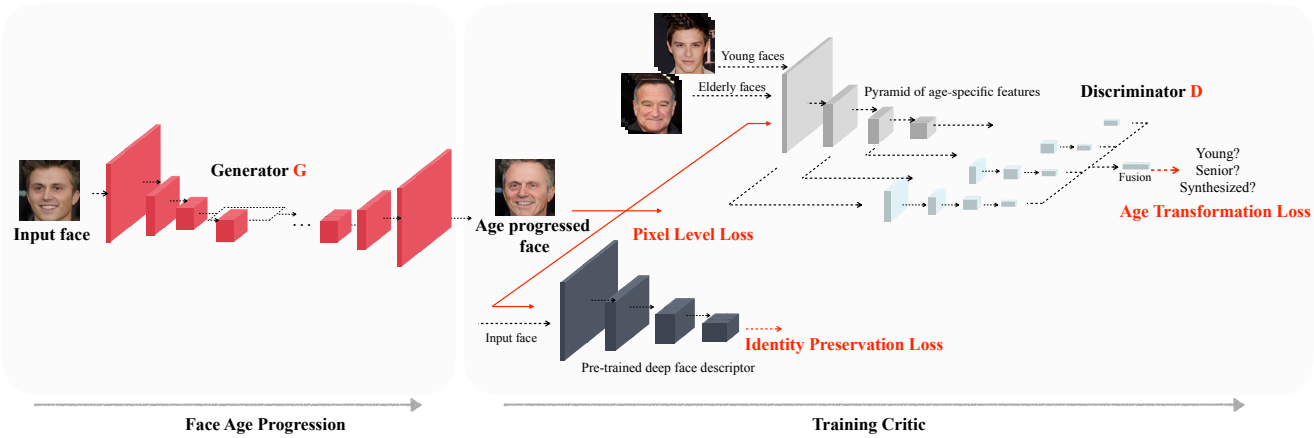

Figure 2. Framework of the proposed age progression method. A CNN based generator G learns the age transformation. The training critic incorporates the squared Euclidean loss in the image space, the GAN loss that encourages generated faces to be indistinguishable from the training elderly faces in terms of age, and the identity preservation loss minimizing the input-output distance in a high-level feature representation which embeds the personalized characteristics.

3. Method

3.1. Overview

A classic GAN contains a generator G and a discriminator D, which are iteratively trained via an adversarial process. The generative function G tries to capture the underlying data density and confuse the discriminative function D, while the optimization procedure of D aims to achieve the distinguishability and distinguish the natural face images from the fake ones generated by G. Both G and D can be approximated by neural networks, *e.g.*, Multi-Layer Perceptron (MLP). The risk function of optimizing this minimax two-player game can be written as:

$$\mathcal{V}(D,G) = \min_G \max_D \mathbb{E}_{x\sim P_{data}(x)} \log[D(x)] + \mathbb{E}_{z\sim P_z(z)} \log[1-D(G(z))] \quad (1)$$

where z is a noise sample from a prior probability distribution P_z, and x denotes a real face image following a certain distribution P_{data}. On convergence, the distribution of the synthesized images P_g is equivalent to P_{data}.

Recently, more emphasis has been given to the conditional GANs (cGANs) where the generative model G approximates the dependency of the pre-images (or controlled attributes) and their corresponding targets. cGANs have shown promising results in video prediction [16], text to image synthesis [23], image-to-image translation [10][37], *etc*. In our case, the CNN based generator takes young faces as inputs, and learns a mapping to a domain corresponding to elderly faces. To achieve aging effects while simultaneously keeping person-specific information stable, a compound critic is exploited, which incorporates the traditional squared Euclidean loss in the image space, the GAN loss that encourages generated faces to be indistinguishable from the training elderly faces in terms of age, and the identity loss minimizing the input-output distance in a high-level feature representation which embeds the personalized characteristics. See Fig. 2 for an overview.

3.2. Generator

Synthesizing age progressed faces only requires a forward pass through G. The generative network is a combination of encoder and decoder. With the input young face, it first exploits three strided convolutional layers to encode it to a latent space, capturing the facial properties that tend to be stable w.r.t. the elapsed time, followed by four residual blocks [9] modeling the common structure shared by the input and output faces, similar to the settings in [11]. Age transformation to a target image space is finally achieved by three fractionally-strided convolutional layers, yielding the age progression result conditioned on the given young face. Rather than using the max-pooling and upsampling layers to calculate the feature maps, we employ the 3×3 convolution kernels with stride of 2, ensuring that every pixel contributes and the adjacent pixels transform in a synergistic manner. All the convolutional layers are followed by batch normalization (BN) and ReLU non-linearity activation. Paddings are added to the layers to make the input and output have exactly the same size. The architecture of G is shown in Table 1 (we do not show BN and ReLU activation for brevity).

3.3. Discriminator

The system critic incorporates the prior knowledge of the data density of the faces from the target age cluster, and a discriminative network D is thus introduced, which outputs a scalar $D(x)$ representing the probability that x comes from the data. The distribution of the generated faces P_g (we denote the distribution of young faces as $x \sim P_{young}$,

Table 1. Generator architecture

Layer	conv.	conv.↓	conv.↓	res.	res.	res.	res.	deconv.↑	deconv.↑	deconv.
Kernel	9	3	3	3	3	3	3	3	3	9
Stride	1	2	2	1	1	1	1	2	2	1
Padding	4	1	1	2	2	2	2	1	1	4
Outputs	32	64	128	128	128	128	128	64	32	3

Table 2. Discriminator architecture

Pathway	Input	Layers (denote as: conv - <ouput>; kernel = 4, stride = 2, padding = 1)					
1	512				conv-512	conv-512	conv-1
2	256			conv-512	conv-512	conv-512	conv-1
3	128		conv-256	conv-512	conv-512	conv-512	conv-1
4	64	conv-128	conv-256	conv-512	conv-512	conv-512	conv-1

then $G(x) \sim P_g$) is supposed to be equivalent to the distribution P_{old} when optimality is reached. Supposing that we follow the classic GAN [8], which uses a binary cross entropy classification, and the process of training D amounts to minimizing the loss:

$$\mathcal{L}_{GAN\_D} = -\mathbb{E}_{x \in P_{young}(x)} \log[1 - D(G(x))] - \mathbb{E}_{x \in P_{old}(x)} \log[D(x)] \tag{2}$$

It is always desirable that G and D converge coherently; however, D frequently achieves the distinguishability faster in practice, and feeds back vanishing gradients for G to learn, since the JS divergence is locally saturated. Recent studies, *i.e.* the Wasserstein GAN [5], the Least Squares GAN [15], and the Loss-Sensitive GAN [21], reveal that the most fundamental issue lies in how exactly the distance between sequences of probability distributions is defined. Here, we use the least squares loss substituting for the negative log likelihood objective, which penalizes the samples depending on how close they are to the decision boundary in a metric space, minimizing the Pearson $\mathcal{X}^2$ divergence. Further, to achieve more convincing and vivid age-specific facial details, both the actual young faces and the generated age-progressed faces are fed into D as negative samples while the true elderly images as positive ones. Accordingly, the training process alternately minimizes the following:

$$\begin{aligned}\mathcal{L}_{GAN\_D} = &\frac{1}{2}\mathbb{E}_{x \sim P_{old}(x)}[(D_\omega(\phi_{age}(x)) - 1)^2] \\ &+ \frac{1}{2}\mathbb{E}_{x \sim P_{young}(x)}[D_\omega(\phi_{age}(G(x)))^2 + D_\omega(\phi_{age}(x))^2]\end{aligned} \tag{3}$$

$$\mathcal{L}_{GAN\_G} = \mathbb{E}_{x \sim P_{young}(x)}[(D_\omega(\phi_{age}(G(x))) - 1)^2] \tag{4}$$

Note, in (3) and (4), a function ϕ_{age} bridges G and D, which is specially introduced to extract age-related features conveyed by faces, as Fig. 2 shows. Considering that human faces at diverse age groups share a common configuration and similar texture properties, a feature extractor ϕ_{age} is thus exploited independently of D, and outputs high-level feature representations to make the generated faces more distinguishable from the true elderly faces in terms of age.

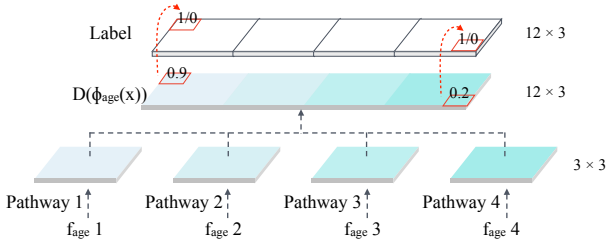

Figure 3. The scores of 4 pathways are finally concatenated and jointly estimated by the discriminator D (D is an estimator rather than a classifier; the *Label* does not need to be a single scalar).

In particular, ϕ_{age} is pre-trained for a multi-label classification task of age estimation with the VGG-16 structure [26], and after convergence, we remove the fully connected layers and integrate it into the framework. Since natural images exhibit multi-scale characteristics, and along the hierarchical architecture, ϕ_{age} captures the properties gradually from exact pixel values to high-level age-specific semantic information, hence this study leverages the intrinsic pyramid hierarchy. The pyramid facial feature representations are jointly estimated by D at multiple scales, handling aging effect generation in a fine-grained way.

The outputs of the 2nd, 4th, 7th and 10th convolutional layers of ϕ_{age} are used. They pass through the pathways of D and are finally concatenated. In D, all convolutional layers are followed by BN and LeakyReLU activation except the last one in each pathway. The detailed architecture of D is shown in Table 2, and the joint estimation on the high-level features is illustrated in Fig. 3.

3.4. Identity Preservation

One core issue of face age progression is keeping the person-dependent properties stable. Therefore, we incorporate the associated constraint by measuring the input-output distance in a proper feature space, which is sensitive to the identity change while relatively robust to other variations. Specifically, the network of *deep face descriptor* [20] is utilized, denoted as ϕ_{id}, to encode the personalized information and further define the identity loss function. ϕ_{id} is

Table 3. Statistics of face aging databases used for evaluation

Database	Number of images	Number of subjects	Number of images per subject	Time lapse per subject (years)	Age span (years old)	Average age (years old)
MORPH [24]	52,099	12,938	1 - 53 (avg. 4.03)	0 - 33 (avg. 1.62)	16 - 77	33.07
CACD [6]	163,446	2,000	22 - 139 (avg. 81.72)	7 - 9 (avg. 8.99)	14 - 62	38.03
FG-NET [4]	1,002	82	6 - 18 (avg. 12.22)	11 - 54 (avg. 27.80)	0 - 69	15.84

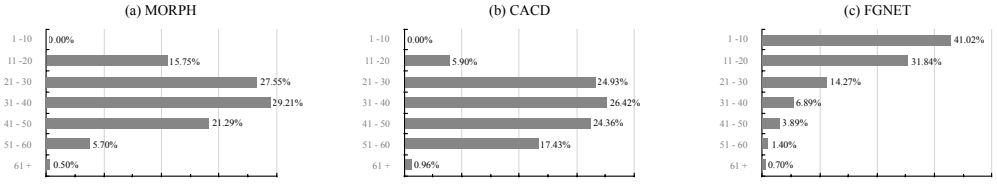

Figure 4. Age distributions of (a) MORPH, (b) CACD, and (c) FGNET.

trained with a large face dataset containing millions of face images from thousands of individuals[1]. It is originally bootstrapped by considering recognizing $N = 2,622$ unique individuals; and then the last classification layer is removed and $\phi_{id}(x)$ is tuned to improve the capability of verification in the Euclidean space using a triplet-loss training scheme. In our case, ϕ_{id} is clipped to have 10 convolutional layers, and the identity loss is then formulated as:

$$\mathcal{L}_{identity} = \mathbb{E}_{x \in P_{young}(x)} d(\phi_{id}(x), \phi_{id}(G(x))) \tag{5}$$

where d is the squared Euclidean distance between feature representations. For more implementation details of *deep face descriptor*, please refer to [20].

3.5. Objective

Besides the specially designed age-related GAN critic and the identity permanence penalty, a pixel-wise L2 loss in the image space is also adopted for further bridging the input-output gap, *e.g.,* the color aberration, which is formulated as:

$$\mathcal{L}_{pixel} = \frac{1}{W \times H \times C} \|G(x) - x\|_2^2 \tag{6}$$

where x denotes the input face and W, H, and C correspond to the image shape.

Finally, the system training loss can be written as:

$$\mathcal{L}_G = \lambda_a \mathcal{L}_{GAN\_G} + \lambda_p \mathcal{L}_{pixel} + \lambda_i \mathcal{L}_{identity} \tag{7}$$

$$\mathcal{L}_D = \mathcal{L}_{GAN\_D} \tag{8}$$

We train G and D alternately until optimality, and finally G learns the desired age transformation and D becomes a reliable estimator.

[1]The face images are collected via the Google Image Search on the names of 5K celebrities, purified by automatic and manual filterings.

4. Experimental Results

4.1. Data Collection

The sources of face images for training GANs are the MORPH mugshot dataset [24] with standardized imaging and the Cross-Age Celebrity Dataset (CACD) [6] involving PIE variations.

MORPH is a large publicly available aging database [24], containing subject's ethnicity, height, weight and gender. An extension of MORPH contains 52,099 color images with near-frontal pose, neutral expression, and uniform illumination (some minor pose and expression variations are indeed present). The subject age ranges from 16 to 77 years old, with the average age being approximately 33 years. The longitudinal age span of one subject varies from 46 days to 33 years. **CACD** is a public dataset [6] collected via the Google Image Search, containing 163,446 face images of 2,000 celebrities across 10 years, with age ranging from 14 to 62. The dataset has the largest amount of images with age changes, showing variations in pose, illumination, expression, *etc.* being less controlled than MORPH. We mainly use MORPH and CACD for training and validation. FG-NET [4] is also adopted for testing to make a fair comparison with prior work, which is popular in face aging analysis but only contains 1,002 images from 82 individuals. See Table 3 and Fig. 4 for more properties of these databases.

4.2. Implementation Details

Prior to feeding the images into the networks, the faces are aligned using the eye locations provided by the dataset itself (CACD) or detected by the online face recognition API of Face++ [3] (MORPH). Excluding those images undetected in MORPH, 163,446 and 51,699 face images from the two datasets are finally adopted, respectively, and they are cropped to 224 $\times$ 224 pixels. Due to the fact that the number of faces older than 60 years old is quite limited in both databases and neither contains images of chil-

Figure 5. Aging effects obtained on the CACD databases for 24 different subjects. The first image in each panel is the original face image and the subsequent 3 images are the age progressed visualizations for that subject in the [31- 40], [41-50] and 50+ age clusters.

dren, we only consider adult aging. We apply age progression on the faces below 30 years old, synthesizing a sequence of age-progressed renderings when they are in their 30s, 40s, and 50s. We follow the time span of 10 years for each age cluster as reported in many previous studies [35][28][36][32][17].

The architectures of the networks G and D are shown in Table 1 and Table 2. For MORPH, the spring constant λ_p, λ_a, and λ_i are set to 0.10, 300.00 and 0.005, respectively; and they are set to 0.20, 750.00 and 0.005 for CACD. At the training stage, we use Adam with the learning rate of 1×10^{-4} and the weight decay factor of 0.5 for every $2,000$ iterations. We (i) update the discriminator at every iteration, (ii) use the age-related and identity-related critics at every generator iteration, and the (iii) pixel-level critic for every 5 generator iterations. The networks are trained with a batch size of 8 for $50,000$ iterations in total, which takes around 8 hours on a GTX 1080Ti GPU.

4.3. Performance Comparison

4.3.1 Experiment I: Age Progression

Five-fold cross validation is conducted. On CACD, each fold contains 400 individuals with nearly 10,079, 8,635, 7,964, and 6,011 face images from the four age clusters of

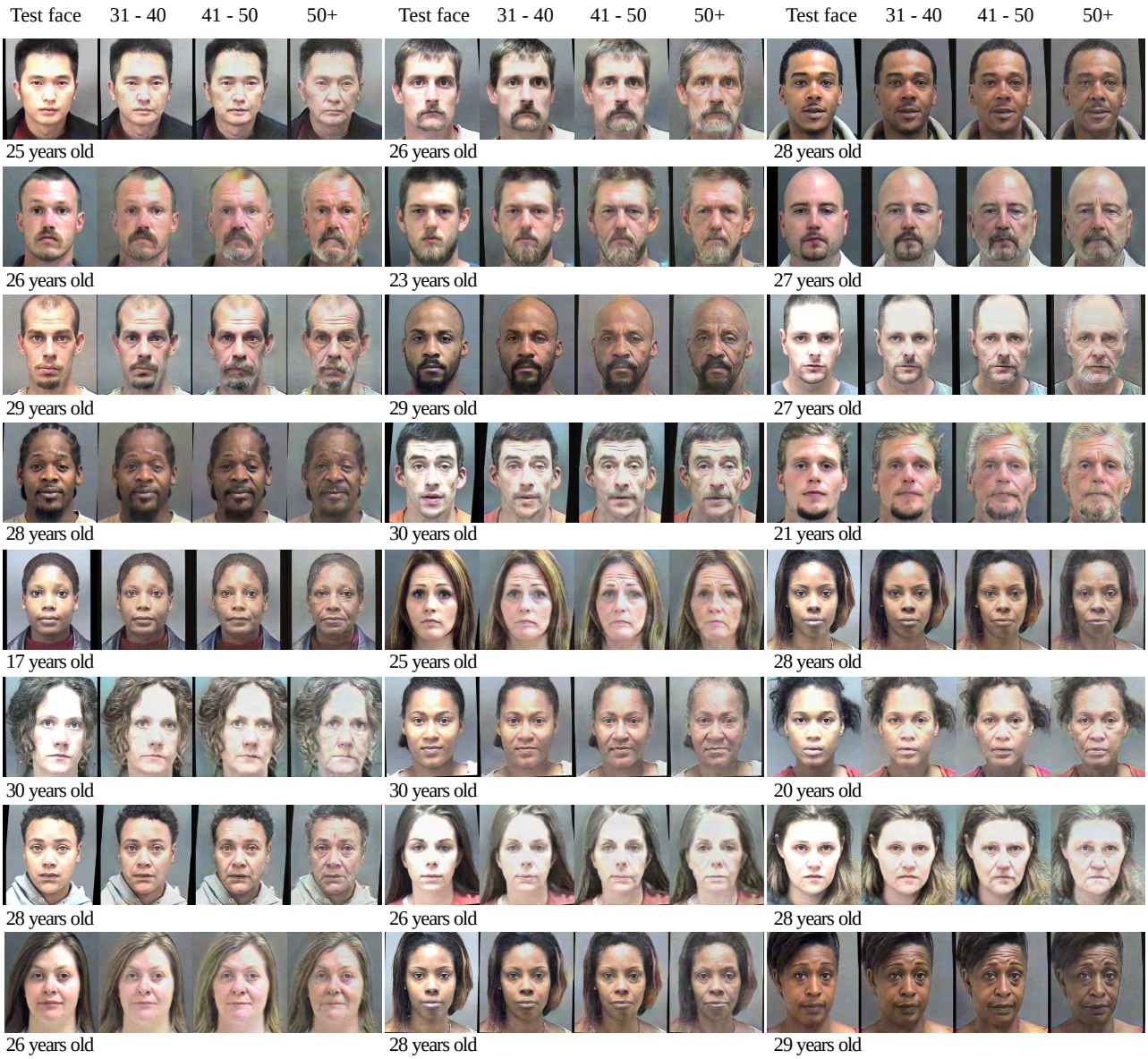

Figure 6. Aging effects obtained on the MORPH databases for 24 different subjects.

[14-30], [31-40], [41-50], and [51-60], respectively; while on MORPH, each fold consists of nearly 2,586 subjects with 4,467, 3,030, 2,205, and 639 faces from the four age groups. For each run, four folds are utilized for training, and the remainder for evaluation. Examples of age progression results are depicted in Fig. 5 and Fig. 6. As we can see, although the examples cover a wide range of population in terms of race, gender, pose, makeup and expression, visually plausible and convincing aging effects are achieved.

The proposed method can also be applied for face rejuvenating simulation. In this experiment, all the test faces come from the people older than 30 years old, and they are transformed to the age bracket of below 30 years old. Example rejuvenating visualizations are shown in Fig. 7. As can be seen, this operation tightens the face skin, and the hair becomes thick and luxuriant as expected.

4.3.2 Experiment II: Aging Model Evaluation

We acknowledge that face age progression is supposed to aesthetically predict the future appearance of the individual, beyond the emerging wrinkles and identity preservation, therefore in this experiment a more comprehensive evaluation of the age progression results are provided with both the visual analysis and the quantitative evaluations.

Experiment II-A: Visual Fidelity: Fig. 8 (a) displays example face images with glasses, occlusions, and pose

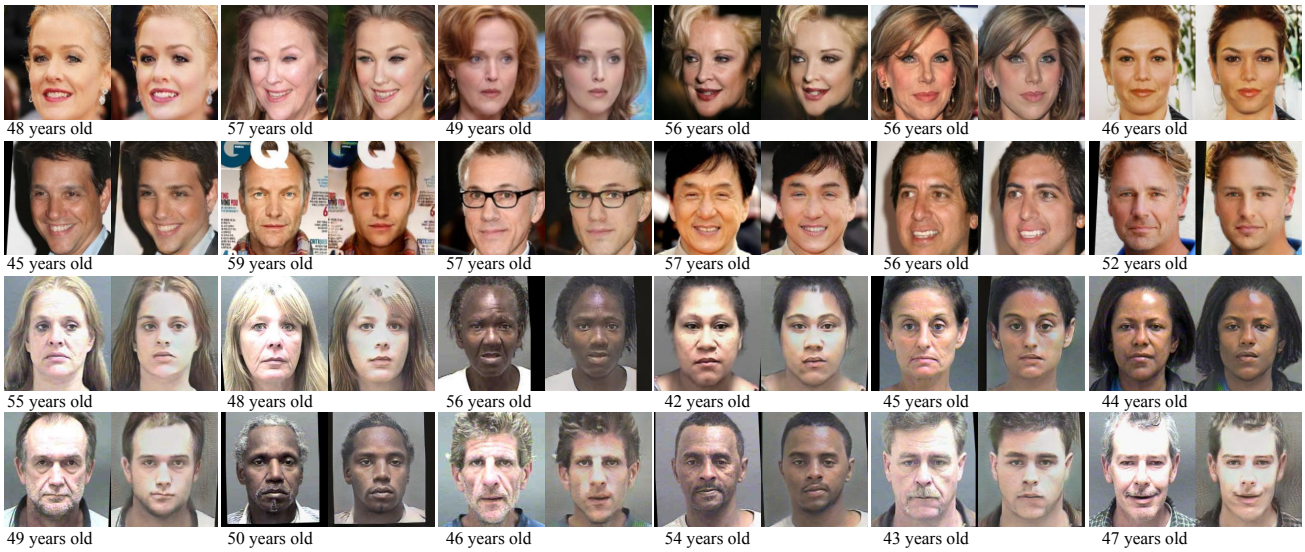

Figure 7. Rejuvenating results achieved on CACD (the top two rows) and MORPH (the bottom two rows) for 24 different subjects. The first image in each panel is the original face image and the second is the corresponding rejuvenating result.

variations. The age-progressed faces are still photorealistic and true to the original inputs; whereas the previous prototyping based methods [28][31] are inherently inadequate for such circumstances, and the parametric aging models [25][27] may also lead to ghosting artifacts. In Fig. 8 (b), some examples of hair aging are demonstrated. As far as we know, almost all aging approaches proposed in the literature [35][25][12][32][36][14] focus on cropped faces without considering hair aging, mainly because hair is not as structured as the face area. Further, hair is diverse in texture, shape, and color, thus difficult to model. Nevertheless, the proposed method takes the whole face as input, and, as expected, the hair grows wispy and thin in aging simulation. Fig. 8 (c) confirms the capability of preserving the necessary facial details during aging, and Fig. 8 (d) shows the smoothness and consistency of the aging changes, where the lips become thiner, the under-eye bags become more and more obvious, and wrinkles are deeper.

Experiment II-B: Aging Accuracy: Along with face aging, the estimated age is supposed to increase. Correspondingly, objective age estimation is conducted to measure the aging accuracy. We apply the online face analysis tool of Face++ [3] to every synthesized face. Excluding those undetected, the age-progressed faces of 22,318 test samples in the MORPH dataset are investigated (average of 4,464 test faces in each run under 5-fold cross validation). Table 4 shows the results. The mean values are 42.84, 50.78, and 59.91 years old for the 3 age clusters, respectively. Ideally, they would be observed in the age range of [31-40],[41-50], and [51-60]. Admittedly, the lifestyle factors may accelerate or slow down the aging rates for the individuals, leading to deviations in the estimated age from

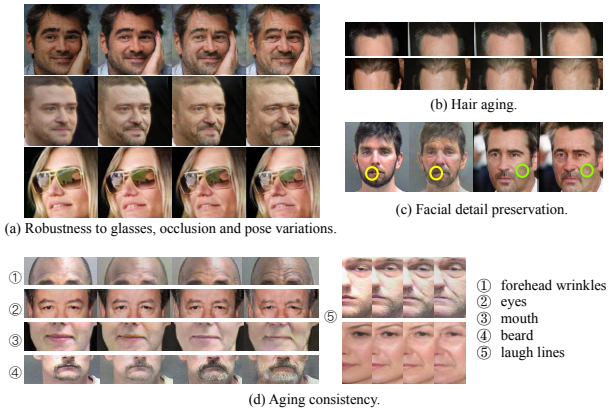

Figure 8. Illustration of visual fidelity (zoom in for a better view).

the actual age, but the overall trends should be relatively robust. Due to such intrinsic ambiguities, objective age estimations are further conducted on all the faces in the dataset as benchmark. In Table 4 and Fig. 9(a), 9(c), it can be seen that the estimated ages of the synthesized faces are well matched with those of the real images, and increase steadily with the elapsed time, clearly validating the method.

On CACD, the aging synthesis results of $50,222$ young faces are used in this evaluation (average of 10,044 test faces in each run). Even though the age distributions of different clusters do not have a good separation as in MORPH, it still suggests that the proposed age progression method has indeed captured the data density of the given subset of faces in terms of age. See Table 4 and Figs. 9(b) and 9(d) for detailed results.

Experiment II-C: Identity Preservation: Objective face verification with Face++ is carried out to check if the

Table 4. Objective age estimation results (in years) on MORPH and CACD

MORPH				CACD			
Age Cluster 0	**Age Cluster 1**	**Age Cluster 2**	**Age Cluster 3**	**Age Cluster 0**	**Age Cluster 1**	**Age Cluster 2**	**Age Cluster 3**
	Synthesized faces*				Synthesized faces*		
–	42.84 ± 8.03	50.78 ± 9.01	59.91 ± 8.95	–	44.29 ± 8.51	48.34 ± 8.32	52.02 ± 9.21
–	42.84 ± 0.40	50.78 ± 0.36	59.91 ± 0.47	–	44.29 ± 0.53	48.34 ± 0.35	52.02 ± 0.19
	Natural faces				Natural faces		
32.57 ± 7.95	42.46 ± 8.23	51.30 ± 9.01	61.39 ± 8.56	38.68 ± 9.50	43.59 ± 9.41	48.12 ± 9.52	52.59 ± 10.48

* The standard deviation in the first row is calculated on all the synthesized faces; the standard deviation in the second row is calculated on the mean values of the 5 folds.

Table 5. Objective face verification results on (a) MORPH and (b) CACD

		Aged 1	**Aged 2**	**Aged 3**		**Aged 1**	**Aged 2**	**Aged 3**
		verification confidence[a]				verification confidence[a]		
	Test face	94.64 ± 0.03	91.46 ± 0.08	85.87 ± 0.25		94.13±0.04	91.96±0.12	88.60±0.15
	Aged 1	–	94.34 ± 0.06	89.92 ± 0.30		–	94.88±0.16	92.63±0.09
	Aged 2	–	–	92.23 ± 0.24		–	–	94.21±0.24
		verification confidence [b]				verification confidence[b]		
(a)	**Test face**	94.64 ± 1.06	91.46 ± 3.65	85.87 ± 5.53	(b)	94.13±1.19	91.96±2.26	88.60±4.19
	Aged 1	–	94.34 ± 1.64	89.92 ± 3.49		–	94.88±0.87	92.63±2.10
	Aged 2	–	–	92.23 ± 2.09		–	–	94.21±1.25
		verification rate (threshold = 76.5, FAR = 1e - 5)				verification rate (threshold = 76.5, FAR = 1e - 5)		
	Test face	100 ± 0 %	98.91 ± 0.40 %	93.09 ± 1.31 %		99.99 ± 0.01 %	99.91 ± 0.05 %	98.28 ± 0.33 %

[a] The standard deviation is calculated on the mean values of the 5 folds.
[b] The standard deviation is calculated on all the synthesized faces.

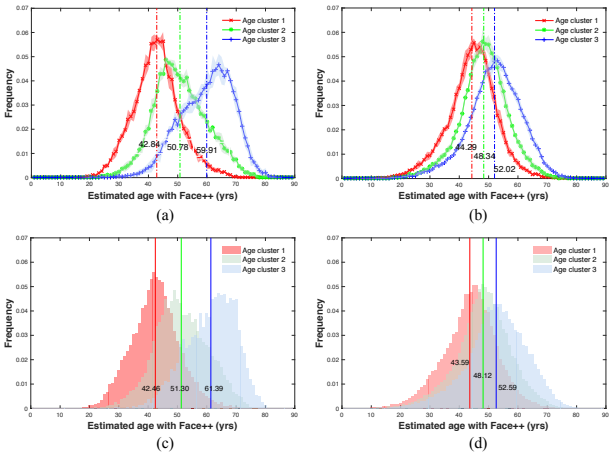

Figure 9. Distributions of the estimated ages obtained by Face++. (a) MORPH, synthesized faces; (b) CACD, synthesized faces; (c) MORPH, actual faces; and (d) CACD, actual faces.

original identity property is well preserved during age progression. For each test face, we perform comparisons between the input image and the corresponding aging simulation results: [test face, aged 1], [test face, aged 2], and [test face, aged 3]; and statistical analyses among the synthesized faces are conducted, *i.e.* [aged 1, aged 2], [aged 1, aged 3], and [aged 2, aged 3]. Similar to Experiment II-B, 22,318 young faces in MORPH and their age-progressed renderings are used in this evaluation, leading to a total of $22,318 \times 6$ verifications. As shown in Table 5, the obtained mean verification rates for the 3 age-progressed clusters are 100%, 98.91%, and 93.09%, respectively. For CACD,

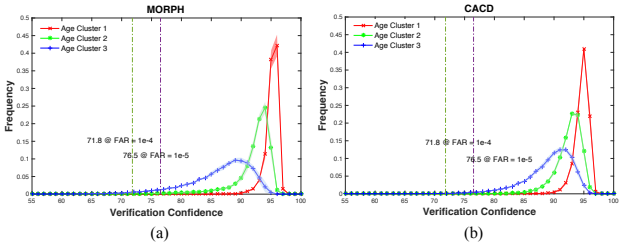

Figure 10. Distributions of the face verification confidence on (a) MORPH and (b) CACD.

there are $50,222 \times 6$ verifications, and the mean verification rates are 99.99%, 99.91%, and 98.28%, respectively, which clearly confirms the ability of identity preservation of the proposed method. Additionally, in Table 5 and Fig. 10, face verification performance decreases as the time elapsed between two images increases, which conforms to the physical truth of face aging, and it may also explain the better performance achieved on CACD compared to MORPH in this evaluation.

Experiment II-D: Contribution of Pyramid Architecture: One model assumption is that the pyramid structure of the discriminator D advances the generation of the aging effects, making the age-progressed faces more natural. Accordingly, we carry out comparison to the one-pathway discriminator, under which scheme the generated faces are directly fed into the estimator rather than represented as feature pyramid first. The discriminator architecture in the contrast experiment is equivalent to a chaining of the network ϕ_{age} and the first pathway in the proposed pyramid D.

Fig. 11 and Fig. 12 provide a demonstration. Visually,

Table 6. Quantitative evaluation results using one-pathway discriminator on (a) MORPH and (b) CACD

		Aged 1	Aged 2	Aged 3
(a)	**Estimated age (yrs old)**	46.14 ± 7.79	54.99 ± 7.08	62.10 ± 6.74
	Verification confidence	93.66 ± 1.15	89.94 ± 2.59	84.51 ± 4.36

		Aged 1	Aged 2	Aged 3
(b)	**Estimated age (yrs old)**	45.89 ± 9.85	51.44 ± 9.78	54.52 ± 10.22
	Verification confidence	92.98 ± 1.76	87.55 ± 4.62	84.61 ± 5.83

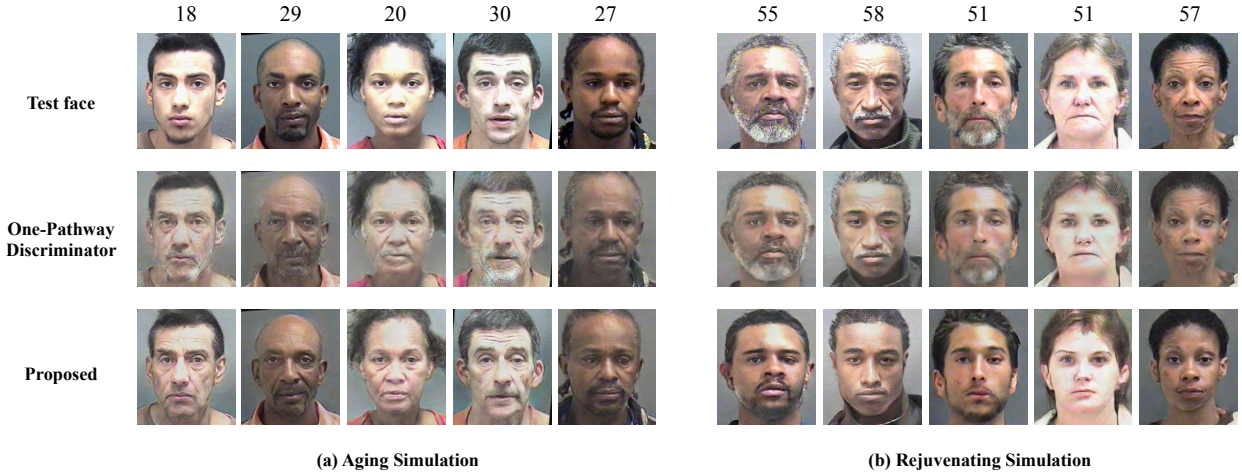

Figure 11. Visual comparison to the one-pathway discriminator on the MORPH database.

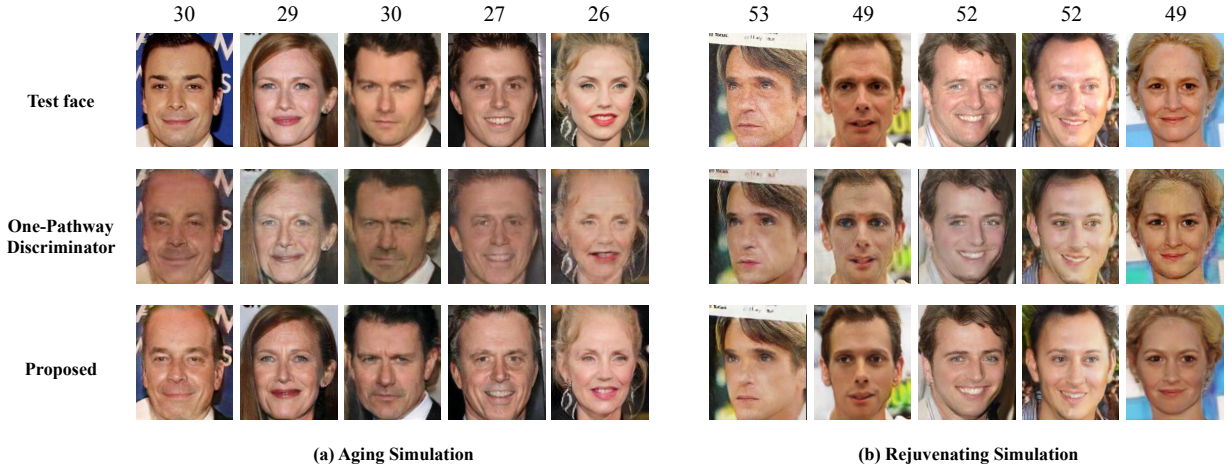

Figure 12. Visual comparison to the one-pathway discriminator on the CACD database.

the synthesized aging details of the counterpart are not so decent. To make the comparison more specific and reliable, quantitative evaluations are further conducted with the similar settings in Experiment II-B and II-C, and the statistical results are shown in Table 6. In the table, the estimated ages achieved on MORPH and CACD are generally older than the benchmark (shown in Table 4), and the mean absolute errors over the three age clusters are 2.69 and 2.52 years for the two databases, respectively, exhibiting larger deviation than 0.79 and 0.50 years obtained by using the pyramid architecture. It is probably because the synthesized wrinkles in this contrast experiment are less neat and the faces look relatively messy. It may also explain the decreased face verification confidence observed in Table 6 in the identity preservation evaluation. Based on both the visual fidelity and the quantitative estimations, we can draw an inference that compared with the pyramid architecture, the one-pathway discriminator, as widely utilized in previous GAN-based frameworks, is lagging behind in regard of modeling the sophisticated aging changes.

Experiment II-E: Comparison to Prior Work: To compare with prior work, we conduct testing on the FG-NET and MORPH databases with CACD as the training set. These studies are [25][27][32][35][18][36][17][19], which signify the state-of-the-art; and moreover, one of the most popular mobile aging applications, *i.e. Agingbooth* [1], and the online aging tool *Face of the future* [2] are also compared. Fig. 13 displays some example faces. As can be seen, *Face of the future* and *Agingbooth* follow the prototyping-based method, where the identical aging mask

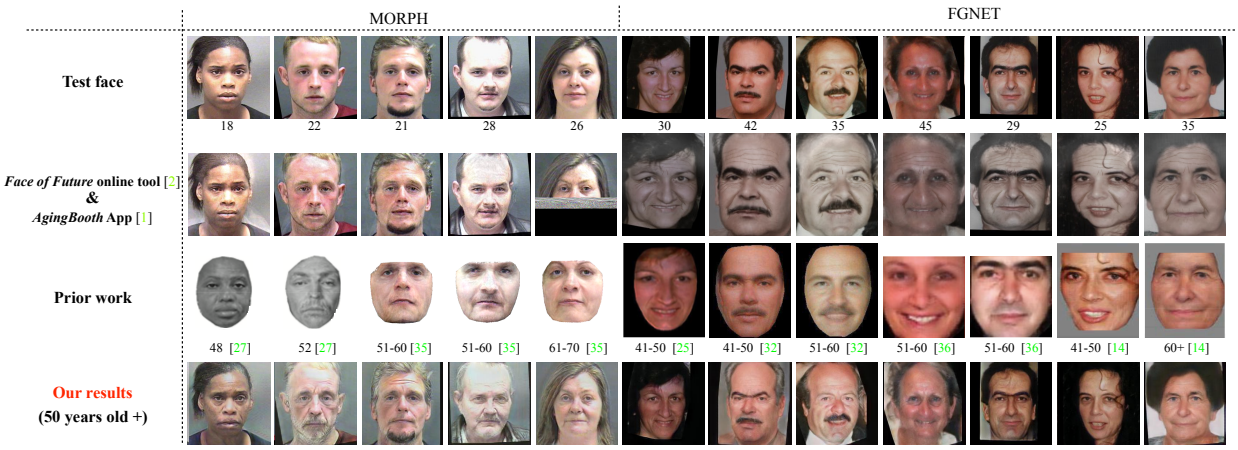

Figure 13. Performance comparison with prior work (zoom in for a better view of the aging details).

is directly applied to all the given faces as most of the aging Apps do. The concept of such methods is straightforward, whereas the age-progressed faces are not photorealistic. Regarding the published works in the literature, ghosting artifacts are ineluctable for the parametric method [27] and the dictionary reconstruction based solution [35][25]. Technological advancements can be observed in the deep generative models [32][36][14], whereas they only focus on the cropped facial area, and the age-progressed faces are short of necessary aging details. In a further statistical survey, we collect 138 paired images of 54 individuals from the published papers, and invite 10 human observers to evaluate which age-progressed face is better in the pairwise comparison. Among the 1380 votes, 69.78% are for ours, 20.80% are for prior work, and 9.42% indicate that they are even. Besides, this proposed method does not require burdensome preprocessing as previous works do, and it only needs 2 landmarks for pupils alignment. To sum up, we can say that the proposed method outperforms the counterparts.

5. Conclusions

Compared with the previous approaches to face age progression, this study shows a different but more effective solution to its key issues, *i.e.* age transformation accuracy and identity preservation, and proposes a novel GAN based method. This method involves the techniques on face verification and age estimation, and exploits a compound training critic that integrates the simple pixel-level penalty, the age-related GAN loss achieving age transformation, and the individual-dependent critic keeping the identity information stable. For generating detailed signs of aging, a pyramidal discriminator is designed to estimate high-level face representations in a finer way. Extensive experiments are conducted, and both the achieved aging imageries and the quantitative evaluations clearly confirm the effectiveness and robustness of the proposed method.

References

[1] *AgingBooth*. PiVi & Co. https://itunes.apple.com/us/app/agingbooth/id357467791?mt=8. 10

[2] *Face of the future*. Computer Science Dept. at Aberystwyth University. http://cherry.dcs.aber.ac.uk/Transformer/index.html. 10

[3] *Face++ Research Toolkit*. Megvii Inc. http://www.faceplusplus.com. 5, 8

[4] *The FG-NET Aging Database*. http://www.fgnet.rsunit.com/ and http://www-prima.inrialpes.fr/FGnet/. 5

[5] M. Arjovsky, S. Chintala, and L. Bottou. *Wasserstein GAN*. arXiv preprint arXiv:1701.07875, 2017. 4

[6] 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 TMM*, 17(6):804–815, 2015. 5

[7] A. Dosovitskiy and T. Brox. Generating images with perceptual similarity metrics based on deep networks. In *NIPS*, 2016. 1

[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NIPS*, pages 2672–2680, 2014. 1, 4

[9] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 2, 3

[10] P. Isola, J. Zhu, T. Zhou, and A. Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004*, 2016. 1, 3

[11] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, pages 694–711. Springer, 2016. 3

[12] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. M. Seitz. Illumination-aware age progression. In *CVPR*, pages 3334–3341, Jun. 2014. 1, 2, 8

[13] A. Lanitis. Evaluating the performance of face-aging algorithms. In *FG*, pages 1–6, 2008. 1

[14] S. Liu, Y. Sun, W. Wang, R. Bao, D. Zhu, and S. Yan. Face aging with contextural generative adversarial nets. In *ACM Multimedia*, 2017. 1, 8, 11

[15] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley. Least squares generative adversarial networks. *arXiv preprint ArXiv:1611.04076*, 2016. 4

[16] M. Mathieu, C. Couprie, and Y. LeCun. Deep multi-scale video prediction beyond mean square error. *arXiv preprint arXiv:1511.05440*, 2015. 3

[17] C. Nhan Duong, K. Gia Quach, K. Luu, N. Le, and M. Savvides. Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition. In *ICCV*, 2017. 1, 2, 6, 10

[18] C. Nhan Duong, K. Luu, K. Gia Quach, and T. D. Bui. Longitudinal face modeling via temporal deep restricted boltzmann machines. In *CVPR*, pages 5772–5780, 2016. 1, 10

[19] U. Park, Y. Tong, and A. K. Jain. Age-invariant face recognition. *IEEE PAMI*, 32(5):947–954, 2010. 1, 10

[20] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *BMVC*, 2015. 4, 5

[21] G. Qi. *Loss-Sensitive Generative Adversarial Networks on Lipschitz Densities*. arXiv preprint arXiv:1701.06264, 2017. 4

[22] N. Ramanathan and R. Chellappa. Modeling shape and textural variations in aging faces. In *FG*, pages 1–8, 2008. 1, 2

[23] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text to image synthesis. In *ICML*, volume 3, 2016. 3

[24] K. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *FG*, pages 341–345, 2006. 5

[25] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan. Personalized age progression with aging dictionary. In *ICCV*, pages 3970–3978, 2015. 1, 8, 10, 11

[26] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 4

[27] J. Suo, X. Chen, S. Shan, W. Gao, and Q. Dai. A concatenational graph evolution aging model. *IEEE PAMI*, 34(11):2083–2096, Nov. 2012. 1, 2, 8, 10, 11

[28] J. Suo, S. C. Zhu, S. Shan, and X. Chen. A compositional and dynamic model for face aging. *IEEE PAMI*, 32(3):385–401, Mar. 2010. 1, 6, 8

[29] Y. Taigman, A. Polyak, and L. Wolf. Unsupervised cross-domain image generation. *arXiv preprint arXiv:1611.02200*, 2016. 1

[30] J. T. Todd, L. S. Mark, R. E. Shaw, and J. B. Pittenger. The perception of human growth. *Scientific American*, 242(2):132, 1980. 1, 2

[31] J. Wang, Y. Shang, G. Su, and X. Lin. Age simulation for face recognition. In *ICPR*, volume 3, pages 913–916, 2006. 8

[32] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe. Recurrent face aging. In *CVPR*, pages 2378–2386, 2016. 1, 2, 6, 8, 10, 11

[33] Y. Wang, Z. Zhang, W. Li, and F. Jiang. Combining tensor space analysis and active appearance models for aging effect simulation on face images. *IEEE TSMC-B*, 42(4):1107–1118, Aug. 2012. 1, 2

[34] Y. Wu, N. M. Thalmann, and D. Thalmann. A plastic-visco-elastic model for wrinkles in facial animation and skin aging. In *PG*, pages 201–214, 1994. 1, 2

[35] H. Yang, D. Huang, Y. Wang, H. Wang, and Y. Tang. Face aging effect simulation using hidden factor analysis joint sparse representation. *IEEE TIP*, 25(6):2493–2507, June 2016. 1, 2, 6, 8, 10, 11

[36] Z. Zhang, Y. Song, and H. Qi. Age progression/regression by conditional adversarial autoencoder. In *CVPR*, 2017. 1, 2, 6, 8, 10, 11

[37] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017. 3