Facial Age Synthesis with Label Distribution-Guided Generative Adversarial Network

Facial Age Synthesis with Label Distribution-guided Generative Adversarial Network

Yunlian Sun, Jinhui Tang, Xiangbo Shu, Zhenan Sun and Massimo Tistarelli

***Abstract*—The existing research work on facial age synthesis has been mostly focused on long-term aging (e.g., over an age span of 10 years or more). In this paper, we employ generative adversarial networks (GANs) as a tool to investigate age synthesis over different age spans. Compared with long-term aging, short-term age synthesis suffers from the reduced amount of available training data, which can severely hinder the model training. We conduct a series of experiments to validate this. To facilitate short-term age synthesis, we further propose label distribution-guided generative adversarial network (ldGAN), where each sample is associated with an age label distribution (ALD) rather than a single age group. Accordingly, each sample can contribute not only to the learning of its own age group but also to neighbouring groups' learning. This is useful when addressing short-term aging to cope with the reduced amount of training data. In addition, unlike one-hot encoding which treats age groups as independent from one another, ldGAN can well capture the correlation among different age groups, so that smooth aging sequences can be achieved. The ALD model is integrated into GAN with a two-step process. Firstly, instead of the traditional one-hot encoding, ALD is applied as the condition of the generator. Secondly, we add a sequence of label distribution learners on top of several multi-scale discriminators, with the aim of minimizing the label distribution learning loss when optimizing both the generator and discriminators. Both qualitative and quantitative evaluations are conducted to assess ldGAN's ability in dealing with two core issues of face aging, i.e., aging effect generation and identity preservation. The obtained experimental results demonstrate the effectiveness of ldGAN in both learning short-term aging patterns and coping with the lack of training data.**

***Index Terms*—Facial age synthesis, generative adversarial networks, label distribution learning.**

I. Introduction

WHAT will Mary look like after 20 years or what did she look like 20 years ago? Age synthesis, which allows to predict the future facial appearance (i.e., age progression) or to reconstruct the past facial appearance (i.e., age regression) of an individual, is one of the most intriguing topics in computer vision, biometrics and computer graphics. It has shown potential in diverse applications including finding lost/wanted persons, face recognition robust to aging variations, entertainment and cosmetic studies to cope with aging. The process of physiological aging causes significant changes in both the shape and texture of human faces, thus making the modelling of face aging a very challenging task. Aesthetically synthesizing faces of a subject at different ages thus becomes an extremely difficult task, even for humans. Even though every human is subject to aging, the associated change in the facial appearance differs for each individual. The aging process occurs slowly. It greatly depends on several intrinsic factors, such as genetics, gender, and ethnicity, and extrinsic factors, such as the lifestyle and the working conditions. The paucity of available labeled data further increases the difficulty of designing an age synthesis model which is able to consistently generalize the aging process for different subjects. Existing databases either include very few time distant face images for each individual (shallow) or the available data comes from very few subjects (narrow). Collecting a deep and broad database is crucial but very hard to accomplish in practice.

Nevertheless, automatic facial age synthesis has received tremendous attention, with several efforts devoted to modeling the longitudinal aging process [1], [2]. Early attempts generally investigated the biological structure and aging process of facial features such as the muscles, skin and cranium [3], [4], [5], [6], [7], [8], [9], [10]. These approaches, however, suffer from the complexity of the model and are computationally expensive. They usually require long sequences of personal face images to cover a long time span. In contrast, prototype approaches do not require long face sequences of the same individual [11], [12], [13], [14], [15]. These approaches first divide all available faces into several age groups, then compute a prototype for each group. A test face can be transformed into an age-progressed one by adding the difference between prototypes of two age groups. Even though achieving promising results, prototype methods fail to keep personalized aging patterns [16]. Despite of the success of deep learning in different areas, only recently researchers have turned their attention to synthesizing faces by using deep neural networks (DNN) [17], [18], [19]. For example, in [17] a recurrent neural network was exploited to model the aging pattern between neighbouring age groups. Duong et al. instead investigated temporal restricted boltzmann machines for learning aging transformations [18]. Among various deep models, generative adversarial networks (GANs) have in particular attracted considerable interest and produced impressive results [20], [21], [22], [23], [24], [25],

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2016YFB1001001), in part by the National Natural Science Foundation of China (Grant No. 61603391, 61925204, 61702265 and 61427811) and in part by grants of the Italian Ministry of Research (PRIN 2015 and SPADA). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. William Schwartz. *(Corresponding author: Jinhui Tang.)*

Y. Sun, J. Tang and X. Shu are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, 210094, Nanjing, China. e-mail: {yunlian.sun, jinhuitang, shuxb}@njust.edu.cn

Z. Sun is with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing 100190, China. e-mail: znsun@nlpr.ia.ac.cn

M. Tistarelli is with the Department of Sciences and Information Technology, University of Sassari, Sassari 07100, Italy. e-mail: tista@uniss.it

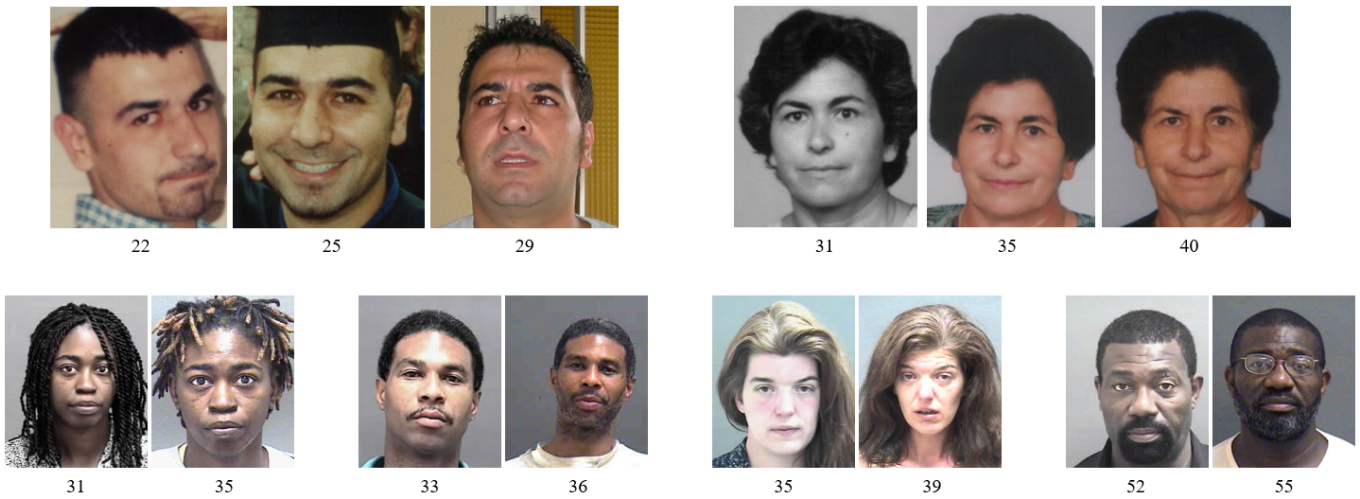

Fig. 1. Visually perceivable changes in facial appearance of six individuals. (Top) Face images captured at different ages of 2 subjects from FG-NET [27]. (Bottom) Face images captured at different ages of 4 subjects from MORPH [28]. The number on the bottom of each face image represents the age of the subject at which the image was captured.

[26].

It should be noted that the past research efforts were mostly devoted to long-term facial age synthesis. For example, all recent approaches proposed for adult face aging consider time spans of 10 years [17], [18], [19], [20], [21], [22], [23], [24], [25], [26] and the corresponding time span for young face aging is 5 years [17], [19], [22]. On the other hand, the analysis of short-term face aging (e.g., over a time frame of less than 5 years) has been relatively understudied. Assuming to adopt a time span of 10 years, which age group model, between $20 \sim 30$ and $30 \sim 40$, should be applied to forecast the facial appearance of a teenager at the age of 30? It is worth noting that both models use data with age differing by even 10 years from 30 for training. In Figure 1, an example of facial age progression of 6 subjects is shown. As depicted in the figure, the pictures of all subjects have been taken in a maximum time frame of 9 years. By adopting a long-term aging model with a span of 10 years, all pictures of the same subject would be classified within the same age group. In this way all shape or texture changes, which are clearly visible in the pictures, would be neglected. On the contrary, by adopting a short-term aging model, pictures of the same subjects would be classified within different age groups, thus allowing to appreciate the characteristic changes of visual features. In general, a short-term aging model allows to capture and better understand even subtle changes in the facial appearance. However, reducing the age span also involves a reduction of the available training data for each age range/group. This side effect should be taken into account when designing short-term aging models.

In this study, we investigate both long-term and short-term facial age synthesis. Specifically, we exploit state-of-the-art conditional generative adversarial networks (cGANs) to generate face aging sequences over different age spans. Experimental results show that GAN can achieve smooth aging sequences and generate high-quality images, when a large age span is used. However, it fails to produce satisfactory results when a small span is used, with many synthesized faces presenting low quality, blurry regions and artifacts.

To facilitate short-term age synthesis, we further propose a novel GAN-based approach, i.e., label distribution-guided generative adversarial network (ldGAN). In ldGAN, each sample is associated with an age label distribution (ALD) rather than a single age group. The label distribution covers different age groups. For each sample, values in its ALD represent the degree that the sample belongs to each age group. Consequently, each sample can contribute not only to the learning of its own age group but also to neighbouring groups' learning. This allows ldGAN to cope with the paucity of training data in short-term age synthesis. Moreover, unlike one-hot encoding which treats age groups as independent from one another, ALD can well capture the correlation among different age groups, so that aging patterns can be well learned. The age label distribution is integrated into the GAN network in two steps. Firstly, ALD is exploited to condition the generator. Secondly, a sequence of label distribution learners are added on top of several multi-scale discriminators. The presented experimental results well demonstrate the effectiveness of ldGAN in both capturing short-term aging patterns and coping with the paucity of training data.

The main contributions of this paper are:

1) We exploit state-of-the-art conditional generative adversarial networks to generate face aging sequences over different age spans.

2) We propose label distribution-guided generative adversarial networks for short-term aging, attempting to fully utilize the limited training data and well capture the correlation among different age ranges.

3) We conduct both qualitative and quantitative experiments to examine ldGAN's ability in both learning aging patterns and keeping identity cues.

The rest of the paper is organized as follows: Section II provides an overview of the related research work. We investigate face aging with different age spans in Section III. Our proposed ldGAN is detailed in Section IV. Section V is

devoted to experimental evaluation. Finally, we conclude the whole work and further give some interesting future work in Section VI.

II. Related Work

In this section, a short literature review on generative adversarial networks, facial age synthesis and label distribution learning is provided.

A. Generative Adversarial Networks

Generative adversarial networks were relatively recently proposed to learn a generative model through an adversarial process [29]. A classical GAN consists of a generator G and a discriminator D. The generator learns to capture the data distribution and generate fake samples that are indistinguishable from real samples. The discriminator instead tries to distinguish real samples from fake ones generated by G. Following this seminal work, various variants have been proposed [30], [31], [32], [33], [34], [35]. The conditional generative adversarial networks have particularly been actively studied [30], where the generator and discriminator are conditioned on some extra information. For example, Tran et al. adopted a pose label as a condition for generating face images with target poses [36]. In order to generate face images with target expression, Song et al. proposed to use facial geometry as the condition of the generator [37]. In StarGAN [38], a target domain label was used as the condition for multi-domain image-to-image translation. In [39], for synthesizing faces with a wide range of expressions, Pumarola et al. introduced a novel GAN conditioning scheme based on Action Units annotations. Similarly, our approach is a conditional GAN, where age label distribution is employed as the condition.

B. Facial Age Synthesis with cGANs

Apart from facial pose synthesis, expression synthesis and attribute manipulation, cGANs have been successfully used in facial age synthesis. In [20], Antipov et al. introduced cGANs to automatic face aging, where a one-hot age group label was used as a condition to guide the synthesis. They achieved promising results for long-term face aging. In [22], a conditional adversarial autoencoder (CAAE) was proposed, where the input face is first encoded to a latent vector. After that the latent vector together with a target one-hot age group label were sent to the generator for age progression/regression. To ensure that the generated faces present desired aging effect, Yang et al. developed an age-related GAN loss for age transformation [23]. They further adopted an identity preservation loss to well keep identity cues during age transformation. Impressive results were obtained for both age progression and rejuvenation. Instead of using an age-related GAN loss, Wang et al. pre-trained an age classifier and used it to determine which age group the face comes from [24]. With this age classification module, the proposed identity-preserved conditional generative adversarial networks (IPCGANs) achieve impressive aging effect. Given that people from different demographic groups have different aging processes, Liu et al. designed an attribute-aware face aging model [26]. Specifically, gender and race attributes were considered. Although achieving impressive results, these approaches focus only on long-term face aging. In this work, we investigate cGANs to generate both long-term and short-term aging sequences.

C. Label Distribution Learning

Label distribution learning (LDL) has been successfully used in several applications including facial age estimation and head pose estimation [40], [41], [42]. This learning model was demonstrated to be more effective with a reduced number of unbalanced training examples. For example, for facial age estimation, each face image is associated with an age label distribution rather than a single age value. Each label distribution covers several age labels, representing the degree to which each label describes the face. Consequently, each sample can contribute not only to the learning of its own age but also to the learning of its neighbouring ages. In addition, unlike one-hot encoding which treats ages as independent from one another, label distribution can successfully capture the correlation among different ages so that samples with neighbouring ages share more than those further away.

III. Long-Term and Short-Term Age Synthesis

In this section, we investigate facial age synthesis with different age spans by using a state-of-the-art GAN architecture.

A. Methodology

Suppose we have a total of N age ranges/groups with an age span of s. We choose $s = 10, 5, 3$ in this study to cover both long-term and short-term face aging. Following existing GAN-based face aging approaches [20], [21], [22], [24], [25], we adopt one-hot label as age encoding. Given a face image I_o, GAN employs I_o as well as a target one-hot label as the condition to generate another image I_t with the same identity but from a different group. Following [24], [25], we train GAN to not only distinguish between real and generated samples but also determine their age groups. That is, in addition to the adversarial loss, we adopt an age group classification loss for optimizing both G and D. In order to preserve the content of the input, we further adopt a reconstruction loss (i.e., cycle-consistency loss [44]). Finally, in order to reduce unfavorable artifacts, we apply total variation regularization to synthesized faces [43]. Note that both [23] and [24] employ perceptual loss, however, we do not observe significant improvements. We thus do not use identity preserving loss here. For the network architecture, we borrow from state-of-the-art GAN architectures like IPCGANs [24], StarGAN [38] and CycleGAN [44]. The generator contains two stride-2 convolution layers for downsampling, six residual blocks, and two stride-2 transposed convolution layers for upsampling. We use instance normalization followed by ReLU activation in all layers except the last output layer, which uses Tanh. For the discriminator, we adopt PatchGANs and add two output layers on top of it. One is used to differentiate

real images from fake ones. The other is used to perform age group classification. We use Leaky ReLU with a negative slope of 0.01 for all six convolution layers of the discriminator, but apply no feature normalization. We apply this GAN framework to facial age synthesis and name it 1hotGAN.

B. Experimental Data

Compared with prior work which investigates only long-term aging, in this work, we study also short-term facial age synthesis. Note that for short-term aging, using inaccurately labeled data will severely interfere with the model training. Suppose we have a training sample with labeled age differing by 8 years from its actual age. If we use $s = 10$, it may be still placed into the true age group. However, if $s = 3$ is adopted, this sample will be placed into a wrong age range differing by even 3 ranges from its true age range. Therefore, a database with accurately labeled data should be used in order to well study short-term aging. Popular facial age databases include MORPH [28], CACD [46], UTKFace [22], FG-NET [27] and Adience [47] databases. Note that ages of CACD are estimated by simply subtracting the birth year from the year of which the photo was taken, thus not accurate enough. Ages of UTKFace instead are estimated through an automatic age estimation algorithm and double checked by a human annotator. They are thus not accurate enough, either. Although face photos in FG-NET are labeled with accurate ages, there include only 1,002 photos. For Adience, face images are grouped into 8 age groups and only group labels are given. The MORPH database is a widely used benchmark for age-related applications, containing a large number of face photos labeled with true ages. We thus perform experiments on MORPH throughout this work. It should be noted that MORPH includes only faces of individuals with ages from 16 to 77 years old. Childhood aging thus cannot be studied with this database, where large transitions occur in face size and structure.

The MORPH database provides not only the real age information but also subjects' gender and ethnicity. We use an extension of this database, which contains 52,099 color images with near-frontal pose, neutral expression, and uniform illumination. Subject ages range from 16 to 77 years old. Since there are very few people who are elder than 50, we do not consider all the ages. For $s = 10$, we define 4 age groups containing 51,158 images, i.e., $16 \sim 25$, $26 \sim 35$, $36 \sim 45$ and $46 \sim 55$. And for $s = 5$, there are 7 groups containing 48,868 images, i.e., $16 \sim 20$, $21 \sim 25$, $26 \sim 30$, $31 \sim 35$, $36 \sim 40$, $41 \sim 45$ and $46 \sim 50$. Finally, we have 12 groups for $s = 3$ containing 49,483 images, i.e., $16 \sim 18$, $19 \sim 21$, $22 \sim 24$, $25 \sim 27$, $28 \sim 30$, $31 \sim 33$, $34 \sim 36$, $37 \sim 39$, $40 \sim 42$, $43 \sim 45$, $46 \sim 48$, and $49 \sim 51$.

For each experiment of $s = 10, 5, 3$, we randomly select around 80% for training and use the remaining for test. Note that subjects in the test set are disjoint from those in the training set. Tables I and II list the data configuration of using different age spans. As observed, when using $s = 10$, each age group has adequate training data. Along with the decrease of s, training data of each group gets less and less. When s reaches to 3, there is only limited training data left for each group. This is particularly the case for Group $49 \sim 51$, with only 1,600 training samples.

For images in MORPH, we first use the tool developed in [48] to detect facial landmarks. Then we perform face alignment using three landmarks, i.e., left eye center, right eye center, and mouth center. The final images are of size $128 \times 128 \times 3$. Although using a larger image size can get more promising results, the model learning process will become very expensive.

C. Synthesis Results

We show in Figure 2 several synthesis results obtained by using 1hotGAN. As can be seen, when $s = 10$ is adopted, 1hotGAN achieves smooth aging sequences and generates high-quality images. When s is decreased to 5, synthesized sequences present a little lower quality. When s is further decreased to 3, generated images show much lower quality. Many synthesized faces present blurry regions and artifacts. Some even present a changed identity. The limited training data in each age group severely hinders the model training process.

IV. Label Distribution-guided GAN

Next, we put emphasis on short-term facial age synthesis, including both age progression and age regression. Specifically, we concentrate on synthesis over an age span of 3 years. As shown in Table II, when $s = 3$ is used, there is very limited training data for each age group. To fully utilize the limited data, we propose to integrate label distribution learning into common GAN architectures. The use of age label distribution further enables our approach to well capture the correlation among different age groups, so that smooth aging sequences can be achieved.

The proposed ldGAN consists of a generator G and 3 multi-scale discriminators denoted as D_1, D_2 and D_3. We use age label distribution as the condition to guide the generator for synthesizing faces with target age groups. The 3 discriminators share the same network architecture but operate at different image scales. Specifically, we downsample the input with a factor of 2 and 4 and create an image pyramid of 3 scales. We then train the 3 discriminators to distinguish between real faces and generated ones. In addition, D_1, D_2 and D_3 undertake the task of age label distribution learning. We achieve this by adding a sequence of label distribution learners on top of the 3 discriminators. Figure 3 illustrates our approach. In the following, we give full details of ldGAN.

A. Label Distribution of Age Groups

LDL was originally proposed for scalar age value estimation [40], where a label distribution covers a certain number of ages. In order to enable ALD to cover different age groups, we use the mean age of each age group as the representative age. Specifically, for each group we compute a mean age and denote it as y_i, where $i = 1, ..., N$. Note that N is the number of age groups. We use i to denote the i-th group. In this study, we adopt Gaussian label distribution for age encoding. For an

TABLE I
NUMBER OF TRAINING AND TEST IMAGES ON MORPH WITH $s = 10$ AND $s = 5$.

	$s = 10$				$s = 5$						
	[16,25]	[26,35]	[36,45]	[46,55]	[16,20]	[21,25]	[26,30]	[31,35]	[36,40]	[41,45]	[46,50]
# training	12,916	10,990	11,777	5,258	6,592	6,324	4,983	6,007	6,182	5,595	3,400
# test	3,274	2,863	2,845	1,235	1,613	1,661	1,388	1,475	1,553	1,292	803

TABLE II
NUMBER OF TRAINING AND TEST IMAGES ON MORPH WITH $s = 3$.

Age Group	[16,18]	[19,21]	[22,24]	[25,27]	[28,30]	[31,33]	[34,36]	[37,39]	[40,42]	[43,45]	[46,48]	[49,51]
# training	3,998	3,728	3,957	3,414	2,802	3,143	4,321	3,619	3,651	3,050	2,306	1,600
# test	1,049	848	991	1,057	717	855	962	907	853	743	537	375

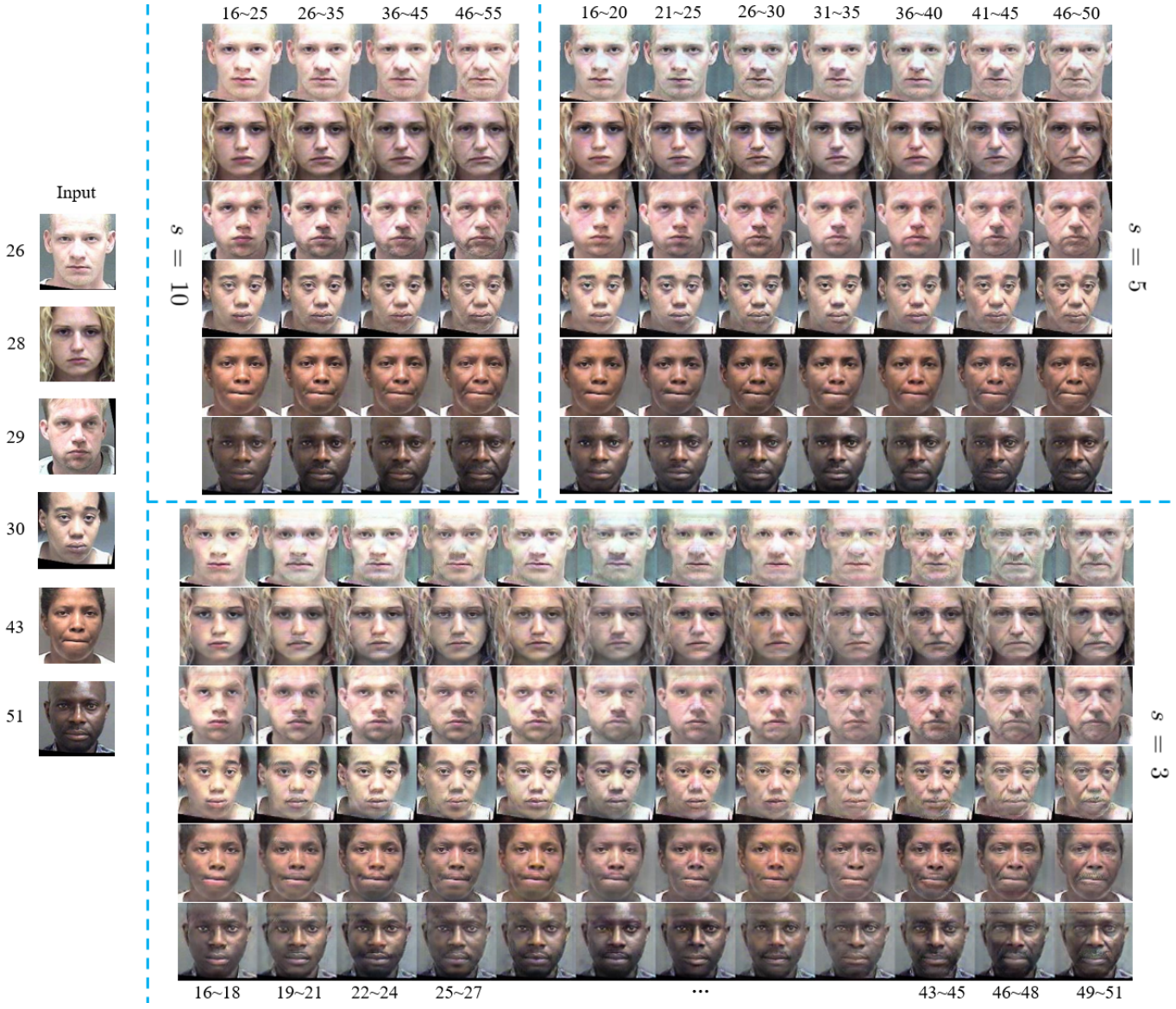

Fig. 2. Face synthesis with different age spans using 1hotGAN. In the leftmost column we give input faces together with their real ages.

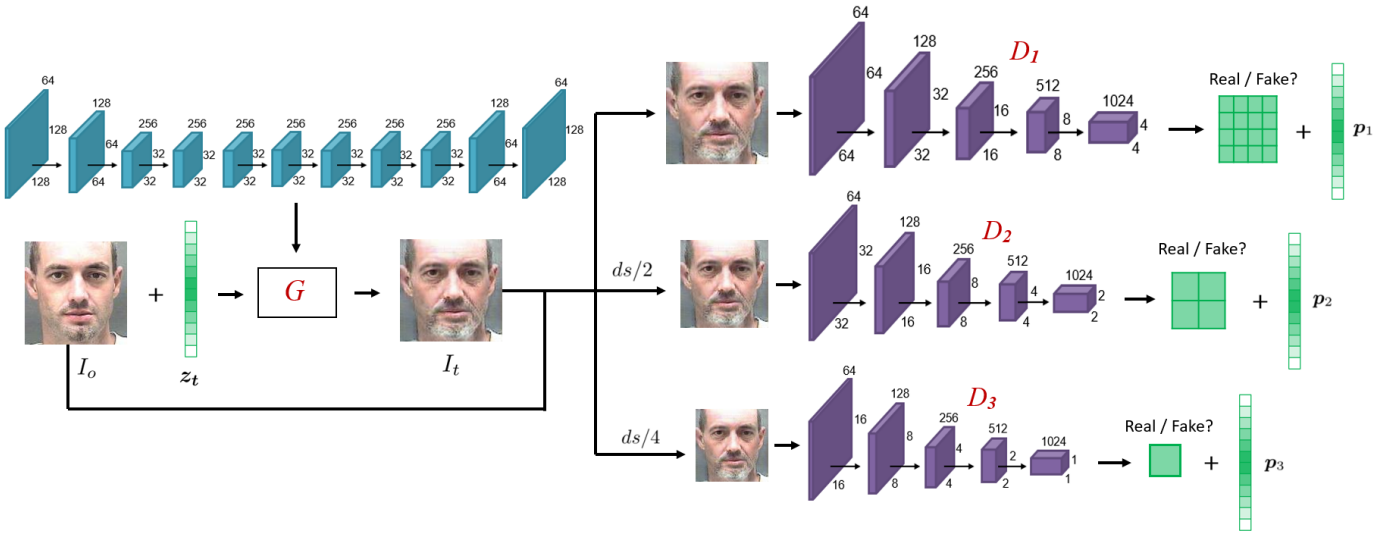

Fig. 3. Framework of ldGAN. We use $ds/2$ and $ds/4$ to denote downsampling operations with factors of 2 and 4. D_1, D_2 and D_3 constitute our multi-scale discriminators. We use $\boldsymbol{p}_1$, $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$ to represent ALDs predicted by the 3 discriminators. Suppose the size of I_o is $H \times W \times 3$, then the input to G has a size of $H \times W \times (3+N)$, where N is the number of age groups.

image from the i-th group, we use y_i as its representative age. The k-th dimension of its ALD is defined as

$$z_k = \frac{1}{\sigma\sqrt{2\pi}W} e^{-\frac{(y_k - y_i)^2}{2\sigma^2}}, k = 1, ..., N, \tag{1}$$

where σ is the standard deviation of Gaussian distribution, and W is the normalization factor which guarantees $\sum_{k=1}^{N} z_k = 1$. That is,

$$W = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=1}^{N} e^{-\frac{(y_k - y_i)^2}{2\sigma^2}}. \tag{2}$$

For a sample from the i-th group, using such an ALD can make sure the description degree of y_i is the highest, while degrees of other groups decrease with the increase of distance away from y_i.

In Table III, we list values of ALD for a sample from the i-th group with different σ and s. We use "-" to denote too tiny values. Note that we list only values of neighbouring groups, since description degrees of further groups are too tiny. As observed, when $s = 10$ is adopted, ALD gets very close to one-hot encoding. This is especially the case for $\sigma = 1$, ALD becomes exactly one-hot encoding. Along with the increase of σ and decrease of s, ALD moves further and further from one-hot encoding and neighbouring groups are given higher and higher values. Following [40], we set $\sigma = 2$ in all our experiments.

TABLE III
VALUES OF NEIGHBOURING GROUPS IN ALD OF A SAMPLE FROM THE i-TH GROUP WITH DIFFERENT σ AND s.

	s	z_i	$z_{i\pm1}$	$z_{i\pm2}$	$z_{i\pm3}$
$\sigma = 1$	10	1.0	-	-	-
	5	0.999995	2.3e-6	-	-
	3	0.9867	0.0066	-	-
$\sigma = 2$	10	0.999993	3.7e-6	-	-
	5	0.9647	0.0177	2.2e-6	-
	3	0.7787	0.1054	0.0052	1.3e-5
$\sigma = 3$	10	0.9923	0.0038	-	-
	5	0.7839	0.1061	0.0019	1.8e-6
	3	0.4949	0.1821	0.0670	0.0033

B. Loss

Given an image I_o with ALD $\boldsymbol{z_o} = \{z_{o1}, ..., z_{oN}\}$, ldGAN aims to synthesize another image I_t with the same identity but from a different age group specified by $\boldsymbol{z_t} = \{z_{t1}, ..., z_{tN}\}$. We randomly generate $\boldsymbol{z_t}$ as the condition of G so that it can flexibly synthesize new images corresponding to different groups. Similarly to 1hotGAN, we use both adversarial and reconstruction loss. However, instead of one-hot encoding based classification loss, we adopt label distribution learning loss when optimizing both the generator and discriminators. Finally, a total variation regularization is applied to synthesized faces with the aim of reducing artifacts.

1) Adversarial Loss: The generator and discriminators are trained alternatively via an adversarial process. Discriminators attempt to distinguish real images from synthesized ones. The generator instead tries to synthesize realistic images that can fool discriminators. ldGAN is conditioned on the input image and a target ALD, the adversarial losses for the generator and discriminators are thus defined as

$$\begin{aligned} L_{adv}^{D} = \frac{1}{3}\sum_{v=1}^{3}\Big\{ &- \mathbb{E}_{I_o}\Big[logD_v(I_o)\Big] - \\ &\mathbb{E}_{I_o,\boldsymbol{z}_t}\Big[log\Big(1 - D_v\big(G(I_o,\boldsymbol{z}_t)\big)\Big)\Big]\Big\}, \end{aligned} \tag{3}$$

$$L_{adv}^{G} = \frac{1}{3}\sum_{v=1}^{3}\mathbb{E}_{I_o,\boldsymbol{z}_t}\Big[log\Big(1 - D_v\big(G(I_o,\boldsymbol{z}_t)\big)\Big)\Big]. \tag{4}$$

2) LDL Loss: Apart from the adversarial loss, our discriminators attempts to minimize LDL loss to ensure both

I_o and I_t are correctly classified into their corresponding age groups. To achieve this, we add a sequence of label distribution learners on top of our multi-scale discriminators. The LDL loss is a cross-entropy loss and is adopted when optimizing both the generator and discriminators. The loss for optimizing discriminators is applied to input images and formulated as

$$L_{ldl}^{D} = \frac{1}{3}\sum_{v=1}^{3}\Big\{\mathbb{E}_{I_o,\boldsymbol{z}_o}\Big[-\sum_{k=1}^{N} z_{ok} log p_{vk}\Big]\Big\}, \tag{5}$$

where $\boldsymbol{p_v} = \{p_{v1}, ..., p_{vN}\}$ is the estimated label distribution of I_o. By minimizing this loss, our discriminators can learn to classify I_o into its corresponding age group specified by $\boldsymbol{z_o}$. On the other hand, the loss used to optimize G is applied to synthesized images and defined as

$$L_{ldl}^{G} = \frac{1}{3}\sum_{v=1}^{3}\Big\{\mathbb{E}_{I_t,\boldsymbol{z}_t}\Big[-\sum_{k=1}^{N} z_{tk} log p_{vk}\Big]\Big\}, \tag{6}$$

where $\boldsymbol{p_v} = \{p_{v1}, ..., p_{vN}\}$ is the learned ALD of I_t. As a result, our generator can learn to generate samples that can be classified into the target age group denoted by $\boldsymbol{z_t}$. With this LDL loss, we can guarantee the aging effect generation.

3) Reconstruction Loss: To ensure the synthesized image preserves the content of its input, we apply a reconstruction loss to G. It takes the form as

$$L_{rec} = \mathbb{E}_{I_o,\boldsymbol{z_t},\boldsymbol{z_o}}[\parallel I_o - G(G(I_o, \boldsymbol{z_t}), \boldsymbol{z_o}) \parallel_1], \tag{7}$$

where G takes in the synthesized image $G(I_o, \boldsymbol{z_t})$ and the original label distribution $\boldsymbol{z_o}$ as input and attempts to reconstruct the original image I_o. We use L_1 norm to encourage less blurring outputs.

4) Overall Objective: Finally, our objective functions to optimize the generator and discriminators are weighted sums of all the above defined losses. They are written, respectively, as

$$L_D = L_{adv}^{D} + \lambda_{ldl} L_{ldl}^{D}, \tag{8}$$

$$L_G = L_{adv}^{G} + \lambda_{ldl} L_{ldl}^{G} + \lambda_{rec} L_{rec} + \lambda_{tv} L_{tv}, \tag{9}$$

where λ_{ldl}, λ_{rec} and λ_{tv} are trade-off parameters. L_{tv} represents the total variation regularization imposed on synthesized samples.

C. Network Architecture

The architecture of the generator keeps the same as 1hot-GAN. The idea of adopting multi-scale discriminators comes from pix2pixHD [45]. For the 3 discriminators, we adopt PatchGANs and add two output layers on each of them. One is used to distinguish between real images and fake ones. It thus outputs the probability of local patches to be real. The other instead implements LDL and outputs estimated ALD. Note that the 3 discriminators share the same network architecture. Similarly to 1hotGAN, we use Leaky ReLU with a negative slope of 0.01 for all convolution layers of the 3 discriminators but apply no feature normalization. Our input to G has the size of $128 \times 128 \times (3 + N)$, while inputs to D_1, D_2 and D_3 are of size $128 \times 128 \times 3$, $64 \times 64 \times 3$ and $32 \times 32 \times 3$, resp.

D. Training Details

Training details keep the same as 1hotGAN. We adopt Wasserstein GAN in order to improve training stability [34]. Our L_{adv}^{D} thus takes the form as

$$L_{adv}^{D} = \frac{1}{3}\sum_{v=1}^{3}\Big\{ -\mathbb{E}_{I_o}\Big[D_v(I_o)\Big] + \mathbb{E}_{I_o,\boldsymbol{z}_t}\Big[D_v\big(G(I_o, \boldsymbol{z}_t)\big)\Big] + \lambda_{gp}\mathbb{E}_{\hat{I}}\Big[\Big(\big\|\nabla_{\hat{I}} D_v(\hat{I})\big\|_2 - 1\Big)^2\Big]\Big\}, \tag{10}$$

where $\hat{I}$ is sampled uniformly along a straight line between I_o and I_t. λ_{gp} is the coefficient of gradient penalty, which is set to be 10 in all our experiments. Trade-off parameters in Eq.9 are set as $\lambda_{ldl} = 4$, $\lambda_{rec} = 10$ and $\lambda_{tv} = 0.0001$. We train ldGAN using Adam with a learning rate of 0.0001, a batch size of 16, $\beta_1 = 0.5$, and $\beta_2 = 0.999$. We perform one optimization step for G after three optimization steps of D_1, D_2 and D_3. Age progression and regression are implemented by randomly generating $\boldsymbol{z}_t$ as the condition of G. It takes about 16 hours to train one model with a single NVIDIA GTX1080Ti GPU. Given a test face image, it costs about 0.04s to generate the whole short-term aging sequence.

V. Experimental Evaluation

In this section, we conduct a series of experiments to evaluate our proposed approach. We first show qualitative results. Then we report objective age estimation and face verification results in order to quantitatively check whether ldGAN performs well in aging effect generation and identity preservation. To thoroughly evaluate ldGAN, we further conduct several ablation studies. We then make a comparison with current state of the art in order to show the superiority of ldGAN over existing related work. Finally, we evaluate the performance of ldGAN in long-term aging scenario, where there is adequate training data. Note that for all the experiments in this section, the data configuration remains unchanged, i.e., keeping the same as Tables I and II.

A. Qualitative Evaluation

Qualitative results are reported in Figure 4. As can been seen, using ldGAN obtains visually plausible results. From young to old, facial skin gets rougher gradually. Nasolabial folds begin to appear. Beard gets white slowly for males. For some people, hair also gets white in old faces. For regression, skin gets smoother gradually. Wrinkles, mustache and beard get reduced or even removed. Apart from the well generated aging effect, we achieve also high-quality images. More importantly, identity is well preserved in synthesized faces.

B. Quantitative Evaluation

To better examine the ldGAN's ability in both learning aging patterns and keeping identity cues, we further perform a quantitative analysis. For evaluating aging effect, we employ the online face analysis tool of Face++ [49] to estimate ages

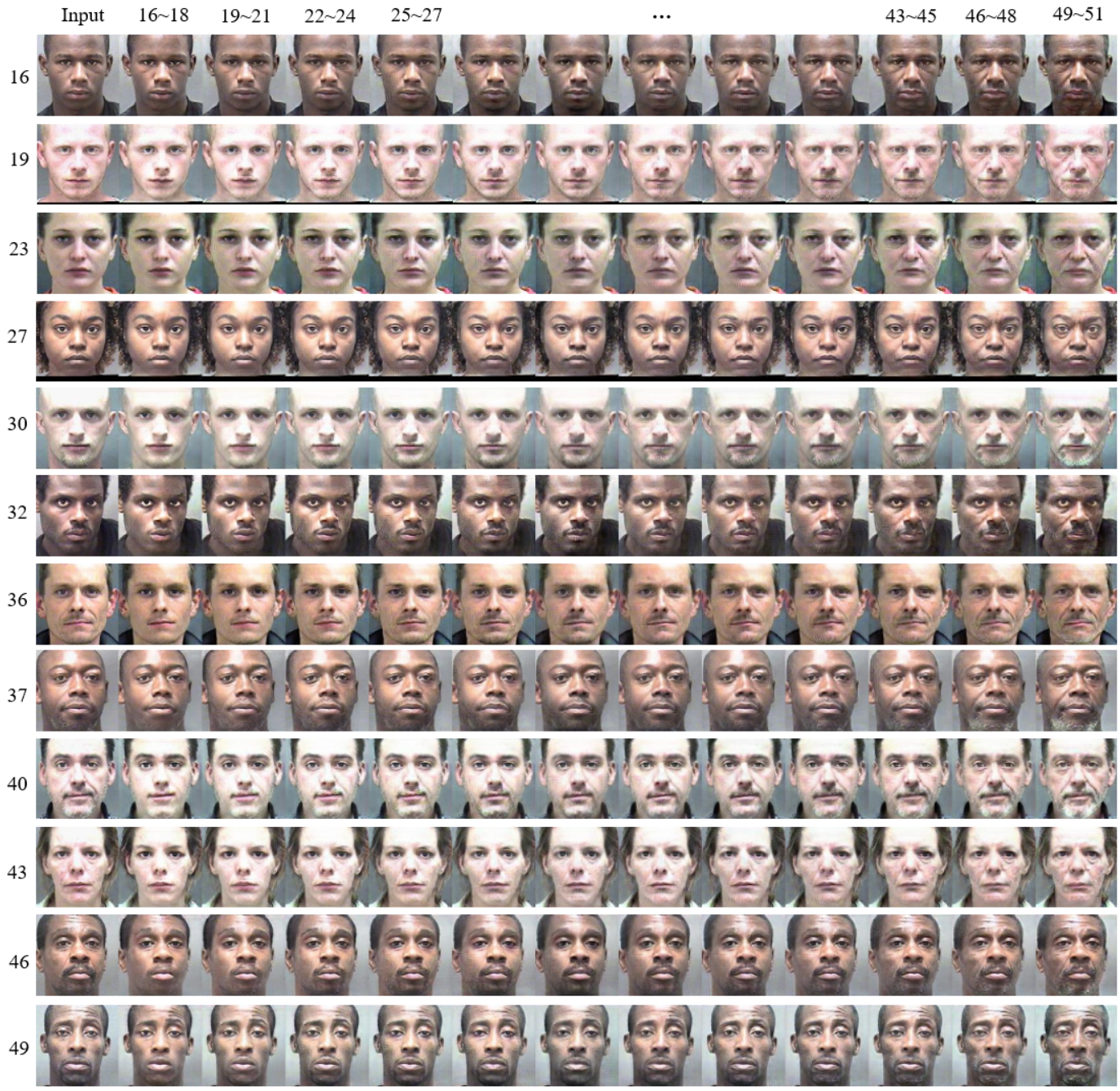

Fig. 4. Qualitative results obtained by using ldGAN. In the leftmost column the real age of each input face is shown.

TABLE IV
OBJECTIVE AGE ESTIMATION RESULTS (IN YEARS) OBTAINED BY FACE++ WITH $s = 3$.

Age Group	$16 \sim 18$	$19 \sim 21$	$22 \sim 24$	$25 \sim 27$	$28 \sim 30$	$31 \sim 33$
Real Face	23.69 ± 4.63	25.64 ± 4.91	28.28 ± 5.32	30.22 ± 5.67	32.76 ± 5.88	35.55 ± 6.43
ldGAN	23.99 ± 4.91	25.30 ± 4.60	26.96 ± 4.69	29.48 ± 5.05	31.64 ± 5.37	35.58 ± 5.63
CAAE	23.88 ± 4.51	25.23 ± 4.87	26.60 ± 5.29	28.04 ± 5.58	29.41 ± 5.91	31.11 ± 6.31
IPCGANs	24.73 ± 5.26	26.52 ± 5.52	28.98 ± 5.81	32.11 ± 6.32	34.69 ± 6.61	35.76 ± 6.63
Age Group	$34 \sim 36$	$37 \sim 39$	$40 \sim 42$	$43 \sim 45$	$46 \sim 48$	$49 \sim 51$
Real Face	37.80 ± 6.77	40.91 ± 7.29	43.41 ± 7.41	46.22 ± 7.59	50.09 ± 7.39	53.15 ± 8.22
ldGAN	40.38 ± 5.76	41.39 ± 6.16	43.25 ± 6.61	47.27 ± 6.43	51.52 ± 6.24	56.52 ± 6.38
CAAE	32.82 ± 6.50	34.23 ± 6.53	36.07 ± 6.89	37.58 ± 7.01	39.18 ± 7.06	40.65 ± 6.94
IPCGANs	39.69 ± 6.86	41.86 ± 6.99	45.44 ± 6.99	49.41 ± 6.64	52.13 ± 6.71	54.39 ± 6.57

TABLE V
OBJECTIVE FACE VERIFICATION RESULTS (CONFIDENCE) OBTAINED BY FACE++ WITH $s = 3$.

Age Group	$16 \sim 18$	$19 \sim 21$	$22 \sim 24$	$25 \sim 27$	$28 \sim 30$	$31 \sim 33$
ldGAN	94.39 ± 1.48	94.69 ± 1.28	94.72 ± 1.13	94.64 ± 1.12	94.47 ± 1.12	94.34 ± 1.17
CAAE	79.92 ± 7.70	80.40 ± 7.42	80.82 ± 7.00	80.91 ± 6.81	81.00 ± 6.59	81.15 ± 6.36
IPCGANs	94.07 ± 1.80	94.77 ± 1.33	95.11 ± 1.04	95.22 ± 0.84	95.21 ± 0.82	95.21 ± 0.82
Age Group	$34 \sim 36$	$37 \sim 39$	$40 \sim 42$	$43 \sim 45$	$46 \sim 48$	$49 \sim 51$
ldGAN	94.67 ± 1.05	94.48 ± 1.06	94.21 ± 1.26	93.91 ± 1.45	93.47 ± 1.66	92.16 ± 2.17
CAAE	81.19 ± 6.22	81.12 ± 6.12	80.93 ± 6.14	80.46 ± 6.35	79.92 ± 6.47	79.55 ± 6.58
IPCGANs	95.05 ± 0.95	94.95 ± 1.06	94.64 ± 1.34	93.82 ± 1.82	93.29 ± 2.19	92.47 ± 2.62

of both test faces (i.e., real faces) and their synthesized faces obtained by ldGAN. For each age group, we calculate the mean and standard deviation of estimated ages. The results are reported in Table IV. Although there exist deviation in Face++'s estimated ages from actual ages, ldGAN keeps well the overall aging trend. By comparing estimated ages of generated faces with those of real faces, we find in most age groups, synthesized faces have very close ages to real faces. Therefore, ldGAN successfully captures short-term aging patterns.

Objective face verification is also conducted using Face++ in order to check whether identity is well preserved during face aging. We compare each test face with its corresponding synthesized faces. A confidence value can then be obtained for each comparison, indicating the similarity of two faces. The confidence lies within [0,100]. Higher confidence indicates higher possibility that two faces are from the same subject. Finally, for each age group, we calculate the mean and standard deviation of confidence over all test faces. The results are shown in Table V. As can be seen, we achieve high confidence for most age groups. Only when synthesizing faces with Group $49 \sim 51$, we obtain a confidence lower than 93. We guess this is caused by the very limited training data, which has only 1,600 samples.

C. Ablation Study

To more completely evaluate our approach, we further present three ablation studies. The first is used to examine the contribution of label distribution learning, which is performed by first replacing ALD with one-hot encoding and then replacing LDL with softmax loss-based age group classification. The second study is performed by using a single-scale discriminator in order to check the contribution of multi-scale discriminators. Finally, we abandon both LDL and multi-scale discriminators, i.e., directly using 1hotGAN investigated in Section III.

The results are given in Figure 5. As observed, without LDL, generated faces show lower quality. Many faces present artifacts and blurry parts. Similar artifacts and blurry parts are also found in faces synthesized by 1hotGAN. When using a single-scale discriminator, there appear beard-like things on female faces. This is especially the case for faces from Groups $28 \sim 30$, $31 \sim 33$ and $40 \sim 42$. ldGAN instead achieves much better results. The use of label distribution learning enables our approach to fully utilize limited training data, so that photorealistic faces can be achieved. And adopting multi-scale discriminators further enhances our label distribution learners' ability.

D. Comparison with Prior Work

Now, we compare our approach with prior work. We select CAAE [22] and IPCGANs [24] as the control methods. On one hand, both approaches adopt one-hot labels as age encoding. On the other hand, their codes are made publicly available on Github, so that we can easily implement both methods for our short-term aging task. We show qualitative results in Figure 6. As observed, CAAE does not perform well in preserving identity. Aging effect is not well generated either. Moreover, faces synthesized by CAAE lack fine details. Compared with CAAE, IPCGANs show superiority in keeping identity cues and generating fine details. However, they are prone to generate old faces with blurred eyes. For synthesis of young male faces, if input faces have beard and mustache, IPCGANs perform poorly when removing them. In contrast, using ldGAN obtains more visually plausible results. We further quantitatively check the ability of CAAE and IPCGANs in aging effect generation and identity preservation. The results are reported in Tables IV and V. As can be seen, CAAE performs poorly in both learning short-term aging patterns and keeping identity cues. IPCGANs show a little higher confidence values than ldGAN in most age groups. However, they perform worse in capturing short-term aging patterns for most age groups. It should be noted that the training of IPCGANs suffers from model collapse. The results of IPCGANs reported here are obtained before model collapse.

E. Generalization Capability Study

Finally, we examine the generalization capability of ldGAN in synthesizing long-term aging sequences with $s = 10$. From Table III, we can see that when $s = 10$ is used, ALD with $\sigma = 2$ gets very close to one-hot encoding. As a result, the contribution of label distribution learning can be omitted. However, we have enough training data for each age group when $s = 10$ is adopted. In Figure 7, we report qualitative results obtained by using ldGAN as well as results achieved by CAAE [22] and IPCGANs [24]. Quantitative results are given in Table VI. As observed, when there is enough training data for each age group, CAAE still performs poorly in both keeping identity cues and learning aging patterns. It fails to generate face images with fine details. By comparison, IPCGANs perform much better in both aging effect generation

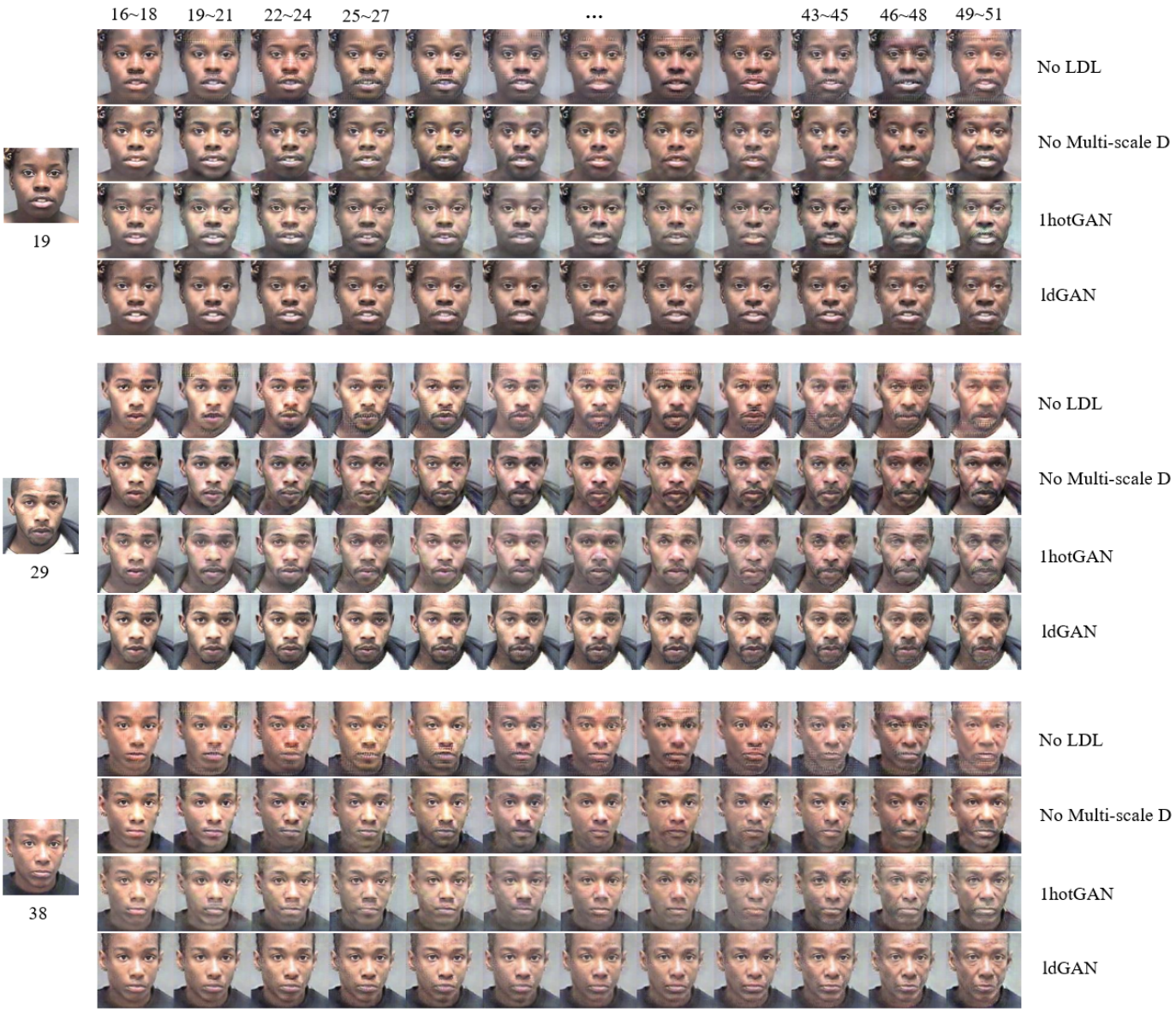

Fig. 5. Ablation study results. Note that 1hotGAN employs neither LDL nor multi-scale D.

and identity preservation. However, when synthesizing images for the last age group, the generation quality gets lower. From Table VI, we can see that ldGAN achieves higher verification confidence than IPCGANs in most age groups. For aging effect comparison, we average the age difference between synthesized and real faces over all age groups. The mean age difference of IPCGANs is 1.02, while ldGAN gets smaller difference 0.63. Note that even with enough training data, IPCGANs still suffers from model collapse. Its results are thus obtained before model collapse. By comparing Table VI with Tables IV and V, we can see that when $s = 10$ is adopted, ldGAN achieves better performance in both learning aging patterns and preserving identity cues. Thus, when there is adequate training data, ldGAN can generate smooth long-term aging sequences.

VI. Conclusion and Future Work

In this paper, we investigated both long-term and short-term facial age synthesis, by employing a state-of-the-art GAN architecture. The presented experimental results showed how a GAN-based model can generate smooth aging sequences with high-quality images, if a large time span is adopted. However, the same model fails to produce satisfactory results when a short time frame is adopted. In order to improve the quality of the face images produced by short-term age synthesis, we proposed a novel GAN-based approach, where each sample is associated with an age label distribution rather than a single age group. The proposed approach works well even when there is a limited amount of training data, owing to the use of ALD. In addition, unlike the one-hot encoding, where age groups are considered as if they were independent from one another, ldGAN can well capture the correlation among different age

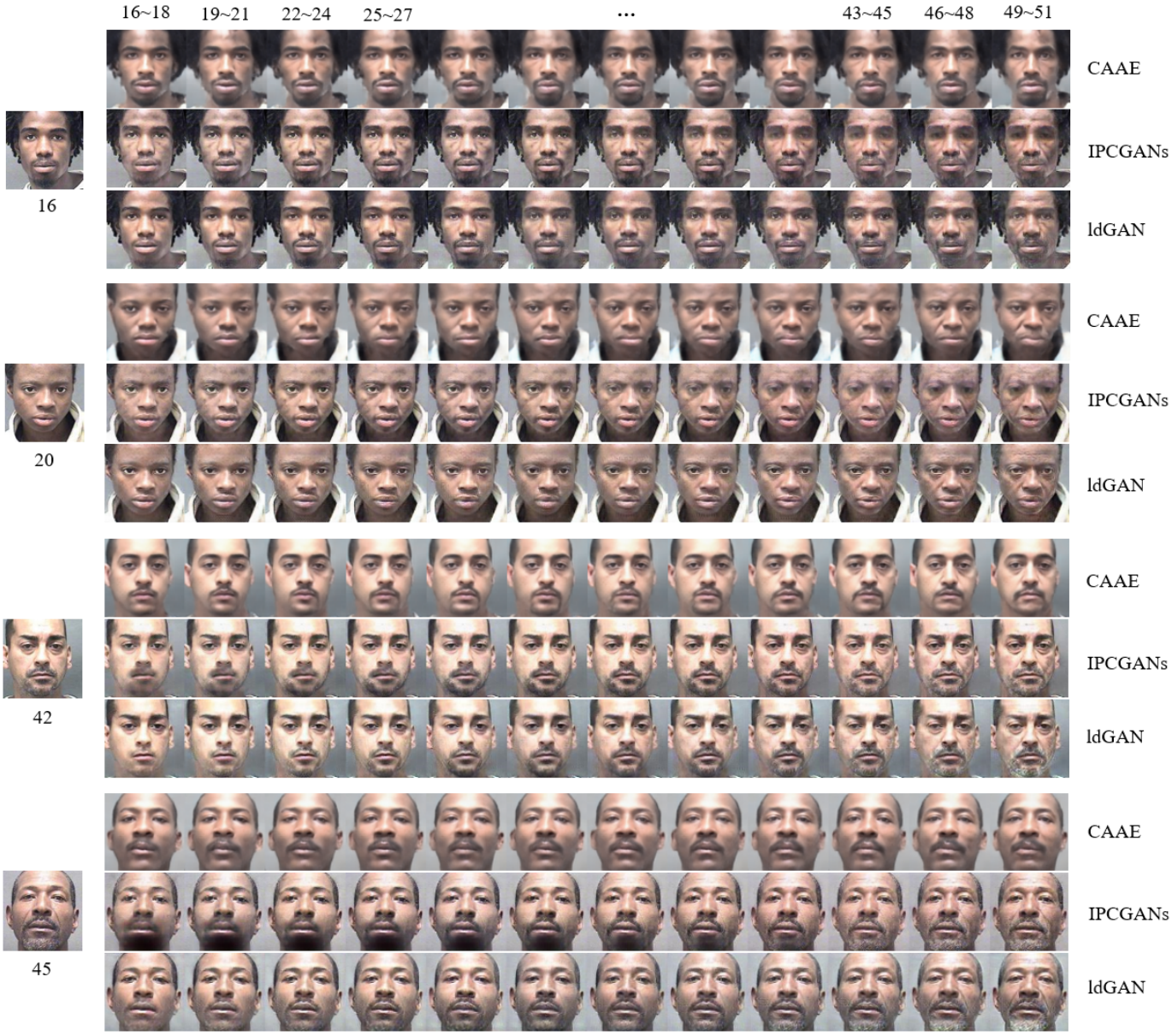

Fig. 6. Comparison with prior work: CAAE [22] and IPCGANs [24]. In the leftmost column we give input faces and their real ages.

TABLE VI
OBJECTIVE AGE ESTIMATION AND FACE VERIFICATION RESULTS OBTAINED BY FACE++ WITH $s = 10$.

	Objective Ages				Verification Confidence			
Age Group	$16 \sim 25$	$26 \sim 35$	$36 \sim 45$	$46 \sim 55$	$16 \sim 25$	$26 \sim 35$	$36 \sim 45$	$46 \sim 55$
Real Face	26.26 ± 5.45	34.05 ± 6.65	42.82 ± 7.76	52.86 ± 8.34	-	-	-	-
ldGAN	25.22 ± 5.67	34.20 ± 6.42	43.84 ± 6.87	53.17 ± 7.57	95.59 ± 0.90	95.53 ± 0.79	95.45 ± 0.83	94.15 ± 1.51
CAAE	23.04 ± 3.96	28.90 ± 5.57	33.07 ± 6.83	40.53 ± 7.39	74.40 ± 9.42	76.09 ± 8.08	76.95 ± 7.40	75.87 ± 7.58
IPCGANs	27.77 ± 7.39	36.18 ± 8.95	43.11 ± 8.93	53.01 ± 7.96	95.46 ± 0.81	95.73 ± 0.61	95.41 ± 0.84	93.07 ± 2.19

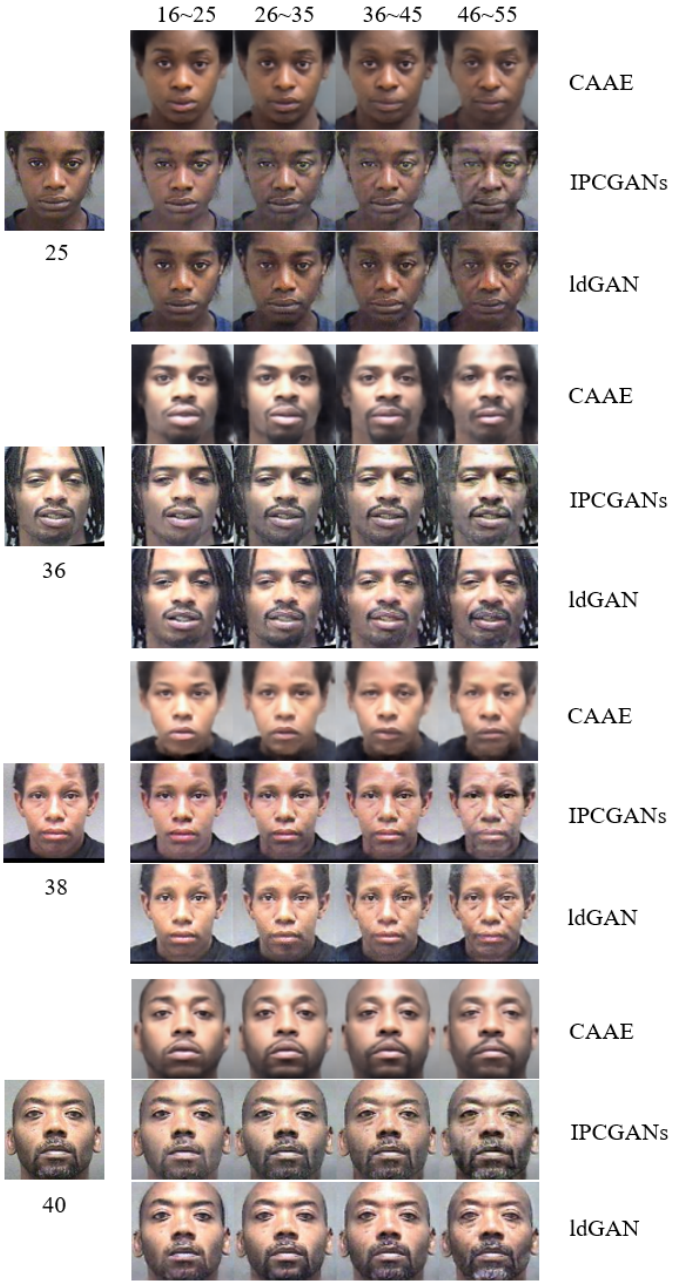

Fig. 7. Long-term aging results with $s = 10$. In the leftmost column we give input faces and their real ages.

groups. The use of multi-scale discriminators further makes the proposed label distribution learners more robust. In order to evaluate ldGAN, both qualitative and quantitative experiments were performed. The obtained results well demonstrated the effectiveness of the proposed approach for both capturing short-term aging patterns and handling the paucity of training data.

Since the MORPH database includes only faces of individuals with ages from 16 to 77 years old, we cannot investigate age progression of children and teenagers with this database. However, facial growth shows large differences from birth to adulthood. Modeling age progression of young faces in a short-term way, thus, will be more significant. In the future, we expect more efforts devoted to both collecting a large number of young faces labeled with accurate ages and studying age progression of the face appearance for children and teenagers in a short-term aging framework.

ACKNOWLEDGMENT

We would like to thank the associate editor and anonymous reviewers for their constructive comments and significant efforts spent to help us for further improving the paper.

REFERENCES

[1] N. Ramanathan and R. Chellappa and S. Biswas, *Computational methods for modeling facial aging: A survey*. Journal of Visual Languages and Computing, vol. 20, no. 3, pp. 131–144, 2009.

[2] Y. Fu and G. Guo and T.S. Huang, *Age Synthesis and Estimation via Faces: A Survey*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 11, pp. 1955–1976, 2010.

[3] N. Ramanathan and R. Chellappa, *Modeling age progression in young faces*. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 387–394, 2006.

[4] N. Ramanathan and R. Chellappa, *Modeling shape and textural variations in aging faces*. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, pp. 1–8, 2008.

[5] A. Lanitis and C.J. Taylor and T.F. Cootes, *Toward Automatic Simulation of Aging Effects on Face Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 442–455, 2002.

[6] J. Suo and S.C. Zhu and S. Shan and X. Chen, *A compositional and dynamic model for face aging*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 3, pp. 385–401, 2010.

[7] Y. Tazoe and H. Gohara and A. Maejima and S. Morishima, *Facial aging simulator considering geometry and patch-tiled texture*. In: Proceedings of the ACM SIGGRAPH Posters, pp. 90, 2012.

[8] Y. Wu and N.M. Thalmann and D. Thalmann, *A dynamic wrinkle model in facial animation and skin ageing*. The journal of visualization and computer animation, vol. 6, no. 4, pp. 195–205, 1995.

[9] Y. Wu and P. Kalra and L. Moccozet and N. Magnenat-Thalmann, *Simulating wrinkles and skin aging*. The visual computer, vol. 15, no. 4, pp. 183–198, 1999.

[10] Z. Liu and Z. Zhang and Y. Shan, *Image-based surface detail transfer*. IEEE Computer Graphics and Applications, vol. 24, no. 3, pp. 30–35, 2004.

[11] D.M. Burt and D.I. Perrett, *Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information*. In: Proceedings of the Royal Society of Londong B: Biological Sciences, pp. 137–143, 1995.

[12] D.A. Rowland and D.I. Perrett, *Manipulating facial appearance through shape and color*. IEEE Computer Graphics and Applications, vol. 15, no. 5, pp. 70–76, 1995.

[13] B. Tiddeman and M. Burt and D. Perrett, *Prototyping and transforming facial textures for perception research*. IEEE Computer Graphics and Applications, vol. 21, no. 5, pp. 42–50, 2001.

[14] Y. Fu and N. Zheng, *M-face: An appearance-based photorealistic model for multiple facial attributes rendering*. IEEE Transactions on Circuits and Systems for Video technology, vol. 16, no. 7, pp. 830–842, 2006.

[15] I. Kemelmacher-Shlizerman and S. Suwajanakorn and S.M. Seitz, *Illumination-aware age progression*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3334–3341, 2014.

[16] X. Shu and J. Tang and H. Lai and L. Liu and S. Yan, *Personalized age progression with aging dictionary*. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3970–3978, 2015.

[17] W. Wang and Z. Cui and Y. Yan and J. Feng and S. Yan and X. Shu and N. Sebe, *Recurrent face aging*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2378–2386, 2016.

[18] C. Nhan Duong and K. Luu and K. Gia Quach and T.D. Bui, *Longitudinal face modeling via temporal deep restricted boltzmann machines*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5772–5780, 2016.

[19] C.N. Duong and K.G. Quach and K. Luu and T.H.N. Le and M. 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, pp. 3755–3763, 2017.

[20] G. Antipov and M. Baccouche and J.L. Dugelay, *Face aging with conditional generative adversarial networks*. In: Proceedings of the IEEE International Conference on Image Processing, pp. 2089–2093, 2017.
[21] S. Liu and Y. Sun and D. Zhu and R. Bao and W. Wang and X. Shu and S. Yan, *Face aging with contextual generative adversarial nets*. In: Proceedings of the 25th ACM international conference on Multimedia, pp. 82–90, 2017.
[22] Z. Zhang and Y. Song and H. Qi, *Age progression/regression by conditional adversarial autoencoder*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[23] H. Yang and D. Huang and Y. Wang and A.K. Jain, *Learning face age progression: A pyramid architecture of gans*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 31–39, 2018.
[24] Z. Wang and X. Tang and W. Luo and S. Gao, *Face aging with identity-preserved conditional generative adversarial networks*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7939–7947, 2018.
[25] P. Li and Y. Hu and R. He and Z. Sun, *Global and Local Consistent Wavelet-domain Age Synthesis*. IEEE Transactions on Information Forensics and Security, 2019.
[26] Y. Liu and Q. Li and Z. Sun, *Attribute-Aware Face Aging With Wavelet-Based Generative Adversarial Networks*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 11877-11886, 2019.
[27] FG-NET. *The FG-NET Aging Database*. http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html, 2002.
[28] K. Ricanek and T. Tesafaye, *Morph: A longitudinal image database of normal adult age-progression*. In: Proceedings of the International Conference on Automatic Face and Gesture Recognition, pp. 341–345, 2006.
[29] I. Goodfellow and J. Pouget-Abadie and M. Mirza and B. Xu and D. Warde-Farley and S. Ozair and A. Courville and Y. Bengio, *Generative adversarial nets*. In: Proceedings of the Advances in Neural Information Processing Systems, pp. 2672–2680, 2014.
[30] M. Mirza and S. Osindero, *Conditional generative adversarial nets*. In: arXiv preprint arXiv:1411.1784, 2014.
[31] A. Radford and L. Metz and S. Chintala, *Unsupervised representation learning with deep convolutional generative adversarial networks*. In: Proceedings of the International Conference on Learning Representations, 2016.
[32] X. Chen and Y. Duan and R. Houthooft and J. Schulman and I. Sutskever and P. Abbeel, *Infogan: Interpretable representation learning by information maximizing generative adversarial nets*. In: Proceedings of the Advances in Neural Information Processing Systems, pp. 2172–2180, 2016.
[33] M. Arjovsky and S. Chintala and L. Bottou, *Wasserstein generative adversarial networks*. In: Proceedings of the International Conference on Machine Learning, pp. 214–223, 2017.
[34] I. Gulrajani and F. Ahmed and M. Arjovsky and V. Dumoulin and A.C. Courville, *Improved training of wasserstein gans*. In: Proceedings of the Advances in Neural Information Processing Systems, pp. 5767–5777, 2017.
[35] X. Mao and Q. Li and H. Xie and R.Y. Lau and Z. Wang and S. Paul Smolley, *Least squares generative adversarial networks*. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2794–2802, 2017.
[36] L.Tran and X. Yin and X. Liu, *Disentangled representation learning gan for pose-invariant face recognition*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7, 2017.
[37] L. Song and Z. Lu and R. He and Z. Sun and T. Tan, *Geometry guided adversarial facial expression synthesis*. In: Proceedings of the ACM Multimedia Conference on Multimedia Conference, pp. 627–635, 2018.
[38] Y. Choi and M. Choi and M. Kim and J.W. Ha and S. Kim and J. Choo, *Stargan: Unified generative adversarial networks for multi-domain image-to-image translation*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[39] A. Pumarola and A. Agudo and A.M. Martinez and A. Sanfeliu and F. Moreno-Noguer, *Ganimation: Anatomically-aware facial animation from a single image*. In: Proceedings of the European Conference on Computer Vision, pp. 818–833, 2018.
[40] X. Geng and C. Yin and Z.H. Zhou, *Facial Age Estimation by Learning from Label Distributions*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 10, pp. 2401–2412, 2013.
[41] Y. Sun and M. Zhang and Z. Sun and T. Tan, *Demographic analysis from biometric data: Achievements, challenges, and new frontiers*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 2, pp. 332–351, 2018.
[42] X. Geng and Y. Xia, *Head pose estimation based on multivariate label distribution*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1837–1842, 2014.
[43] J. Johnson and A. Alahi and F.F. L., *Perceptual losses for real-time style transfer and super-resolution*. In: Proceedings of the European Conference on Computer Vision, pp. 694–711, 2016.
[44] J.Y. Zhu and T. Park and P. Isola and A.A. Efros, *Unpaired image-to-image translation using cycle-consistent adversarial networks*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2242–2251, 2017.
[45] T.C. Wang and M.Y. Liu and J.Y. Zhu and A. Tao and J. Kautz and B. Catanzaro, *High-resolution image synthesis and semantic manipulation with conditional gans*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8798–8807, 2018.
[46] B.C. Chen and C.S. Chen and W.H. Hsu, *Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset*. IEEE Transactions on Multimedia, vol. 17, no. 6, pp. 804–815, 2015.
[47] E. Eidinger and R. Enbar and T. Hassner, *Age and gender estimation of unfiltered faces*. IEEE Transactions on Information Forensics and Security, vol. 9, no. 12, pp. 2170–2179, 2014.
[48] A. Bulat and G. Tzimiropoulos, *How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)*. In: Proceedings of the International Conference on Computer Vision, pp. 4, 2017.
[49] Megvii Inc. *Face++*. https://www.faceplusplus.com/, 2019.

Yunlian Sun received the ME degree in computer science and technology from the Harbin Institute of Technology, China, in 2010 and the Ph.D. degree in ingegneria elettronica, informatica e delle telecomunicazioni from the University of Bologna, Italy, in 2014. She is currently an Associate Professor at the School of Computer Science and Engineering, Nanjing University of Science and Technology, China. Her research interests include biometrics, pattern recognition, and computer vision.

Jinhui Tang received the BEng and Ph.D. degrees from the University of Science and Technology of China, in 2003 and 2008, respectively. He is currently a professor with the Nanjing University of Science and Technology. He has authored more than 150 papers in top-tier journals and conferences. His research interests include multimedia analysis and computer vision. He was a recipient of the best paper awards in ACM MM 2007, PCM 2011 and ICIMCS 2011, the Best Paper Runner-up in ACM MM 2015, and the best student paper awards in MMM 2016 and ICIMCS 2017. He has served as an associate editor of the IEEE TNNLS, the IEEE TKDE, and the IEEE TCSVT. He is a senior member of the IEEE.

Xiangbo Shu received the Ph.D. degree from Nanjing University of Science and Technology in 2016. He is currently an Associate Professor at the School of Computer Science and Engineering, Nanjing University of Science and Technology, China. From 2014 to 2015, he worked as a visiting scholar in the Department of Electrical and Computer Engineering at National University of Singapore. His research interests include computer vision, multimedia computing and deep learning. He has received the Excellent Doctoral Dissertation of CAAI, the Excellent Doctoral Dissertation of Jiangsu Province, the Best Student Paper Award in MMM 2016 and the Best Paper Runner-up in ACM MM 2015. He is a member of the IEEE, ACM, and CCF.

Zhenan Sun received the BE degree in industrial automation from Dalian University of Technology, China, the MS degree in system engineering from Huazhong University of Science and Technology, China, and the Ph.D. degree in pattern recognition and intelligent systems from CASIA, in 1999, 2002, and 2006, respectively. He is currently a professor in the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China. His research interests include biometrics, pattern recognition, and computer vision. He is a fellow of the IAPR.

Massimo Tistarelli received the Ph.D. degree in computer science and robotics from the University of Genoa, Genoa, Italy, in 1991. He is currently a tenured Full Professor of Computer Science and the Director of the Computer Vision Laboratory with the University of Sassari, Sassari, Italy. His main research interests cover biological and artificial vision (in particular, recognition, 3-D reconstruction, and dynamic scene analysis), pattern recognition, biometrics, visual sensors, robotic navigation, and visuomotor coordination. He has coauthored over 100 scientific papers in peer-reviewed books, conferences, and international journals. He is one of the world-recognized leading researchers in biometrics. He is an Associate Editor of IEEE TPAMI, Pattern Recognition Letters, and Image and Vision Computing. He is a fellow of the IAPR.