Face Aging with Identity-Preserved Conditional Generative Adversarial Networks

Zongwei Wang
Shanghaitech University
wangzw@shanghaitech.edu.cn

Xu Tang
Baidu
tangxu02@baidu.com

Weixin Luo, Shenghua Gao*
Shanghaitech University
{luowx, gaoshh}@shanghaitech.edu.cn

Abstract

Face aging is of great importance for cross-age recognition and entertainment related applications. However, the lack of labeled faces of the same person across a long age range makes it challenging. Because of different aging speed of different persons, our face aging approach aims at synthesizing a face whose target age lies in some given age group instead of synthesizing a face with a certain age. By grouping faces with target age together, the objective of face aging is equivalent to transferring aging patterns of faces within the target age group to the face whose aged face is to be synthesized. Meanwhile, the synthesized face should have the same identity with the input face. Thus we propose an Identity-Preserved Conditional Generative Adversarial Networks (IPCGANs) framework, in which a Conditional Generative Adversarial Networks module functions as generating a face that looks realistic and is with the target age, an identity-preserved module preserves the identity information and an age classifier forces the generated face with the target age. Both qualitative and quantitative experiments show that our method can generate more realistic faces in terms of image quality, person identity and age consistency with human observations.

1. Introduction

Face aging, also known as aging synthesis of the human face, is a task of synthesizing faces of a certain person under a given age. It is attracting more and more researchers' attention because of its various applications in cross-age face recognition and entertainment. For example, it could be applied to help find lost children or to predict what someone will look like in the future. Extensive studies have been made on face aging [28] [11] [5]. However, the lack of training samples for a given person over a long range of years [15][20] [3] [21] makes face aging still an extremely challenging task in computer vision.

*Corresponding author.

Traditional face aging methods can roughly be categorized into prototype-based approaches [11] and physical model-based approaches [25]. Prototype-based approaches usually compute an average face within a face age group first, and the difference between different average faces from different groups would be treated as aging pattern which would be used for synthesizing an aged face [11]. Consequently, person-specific information of each person would be lost, which results in the synthesized faces look unrealistic. By contrast, physical model-based approaches model the shape and texture changes with age in terms of hair colors, muscles, and wrinkles, *etc.* with a parametric model, which usually requires lots of training samples and is computationally expensive.

Recently, Generative Adversarial Networks(GANs) based approaches have been demonstrated their successes in generating high quality images [7] [14] [18]. Of which, as a special GANs, Conditional GANs (cGANs) [7] [14] [9] take prior information in image generation and make generated images be with certain desired property. Inspired by the success of CGANs, we propose an Identity-Preserved Conditional Generative Adversarial Networks (IPCGANs) for face aging. Specifically, our IPCGANs consists of three modules: a CGANs module, an identity-preserved module and an age classifier. The generator of CGANs takes an input image and a target age code as its input and generates a face with the target age. The generated face is expected to be indistinguishable from real faces in that target age group by the discriminator. To keep identity information, we introduce a perceptual loss [4] in the objective of IPCGANs. Finally, to guarantee the synthesized faces fall into the target age group, we send the generated aged faces to a pre-trained age classifier and add an age classification loss to the objective. Since all components of our IPCAGANs are differentiable with respect to the model parameters, the whole network can be trained in an end-to-end fashion.

The contributions of this paper are summarized as follows:

1. We propose to impose an identity-preserved term and an age classification term into the objective of our IPC-

GANs. The former lets the aged faces keep the same identity with the input face. The latter is to make sure the generated faces be with the target age. Extensive experiments validate the effectiveness of both terms for preserving the identity information and making the face aging effect evident.

2. Other than quantitatively evaluate the quality of the synthesized faces, we also propose to conduct face verification and face age classification for the generated aged faces by means of user study. Our proposed data augmentation experiment also validates the effectiveness of IPCGANs.

3. IPCGANs is not limited to face aging problem, it is a general framework. Without any modification, IPCGANs can be applied to multi-attribute generation task, like hair colors, facial expressions, etc, which can be used for imbalanced data classification scenes.

2. Related Work

2.1. Face Aging

As aforementioned, traditional face aging approaches can be categorized into prototype-based approaches and physical model based approaches. We refer readers to [5] for a comprehensive survey of these approaches. Specifically, physical model-based approaches usually focus on the change of skin's anatomy structure, facial muscle changes and some other physical measurements for aged face modeling [25][27]. These models are usually very complex, and require lots of training data. Prototype-based approaches leverage the differences between the average faces of different age groups for age pattern transfer [5] [11]. However, such strategy neglects the differences between different persons, which makes the generated faces look unrealistic. Further, some important age clues, say wrinkles, may be averaged out. To avoid this, in [23][30] [29], sparse representation based approaches have been adopted to model the person-specific facial properties for synthesizing aged faces. Though the identity information can be preserved to some extent by these methods, the reconstruction procedure makes the synthesized faces suffer from the ghost artifacts. Recently, a recurrent face aging framework [28] has been proposed for face aging by leveraging a Recurrent Neural Network model. Thus the change of synthesized faces between neighboring age groups is more smooth, but the identity information is not explicitly preserved in this work. [2] is the first to apply conditional GANs to face aging. Their training process is three-stage. This method is not efficient at inference time because they have to solve a LBFGS optimization problem for each image. To better preserve the identity information, they propose a Local Manifold Adaptation approach in [1]. Combined with [2], they boost the cross-age face verification via age normalization. Similar to us, [31] proposed an auto-encoder conditional GANs which encodes the input image to a manifold and then reconstructs aged images. However, their aged faces seem little change given different age conditions. Recently, [24] proposed a face editing method which can be extended to face aging task. Their results show some aging effect, but the aged faces look blurry.

2.2. Generative Adversarial Networks (GANs)

Generative Adversarial Networks(GANs) [7] has been widely used for image generation. It has two components: a generator and a discriminator. Given a noise vector z which is sampled from a normal distribution or a uniform distribution $p_z(z)$, the generator maps z to a synthesized image $\tilde{x}$. The discriminator takes either $\tilde{x}$ or x (x is images sampled from real image distribution $p_{data}(x)$) as input and tries to tell them apart. The generator is trained to let the discriminator be unable to discriminate them. The objective function of original GANs is given as follows:

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

To facilitate the training of GANs, Radford et al. [18] propose a Deep Convolutional GANs (DCGANs) framework, which promotes the application of GANs in many tasks, such as video prediction [13], cross domain image generation [26], *etc*. Arjovsky *et al.* provide a rigorous analysis on the objective of GAN and its instability in training phase, which leads to a Wasserstein GANs (WGANs). Soon after WGANs, an improved version of WGANs is proposed [8]. In [14], a Conditional GANs(CGANs) model which employs prior information in image generation is proposed. Reed *et al.*[19] demonstrate its capability in generating realistic images from text descriptions. Recently, CycleGANs [32] has also been successfully applied to image-to-image translation task and achieves good performance. These work greatly boosts the performance of GANs in image generation.

2.3. Style transfer

The objective of synthesizing a face with a target age is also related to the work of style transfer [10] [6]. Given one input image (to be transferred with some artistic style) and one artistic style image, the goal of style transfer is to generate one image whose contents are taken from the former while the style is from the latter. To reach this goal, a content loss and a style loss in feature space are jointly optimized [10]. Specifically, both the content loss and the style loss are called as perceptual loss because they depend on features extracted from a pre-trained neural network. A neural network extracts more abstract and perceptual meanings features than raw pixel features. While [6] can generate high quality images, it is slow in testing phase because the inference needs to solve a LBFGS optimization problem.

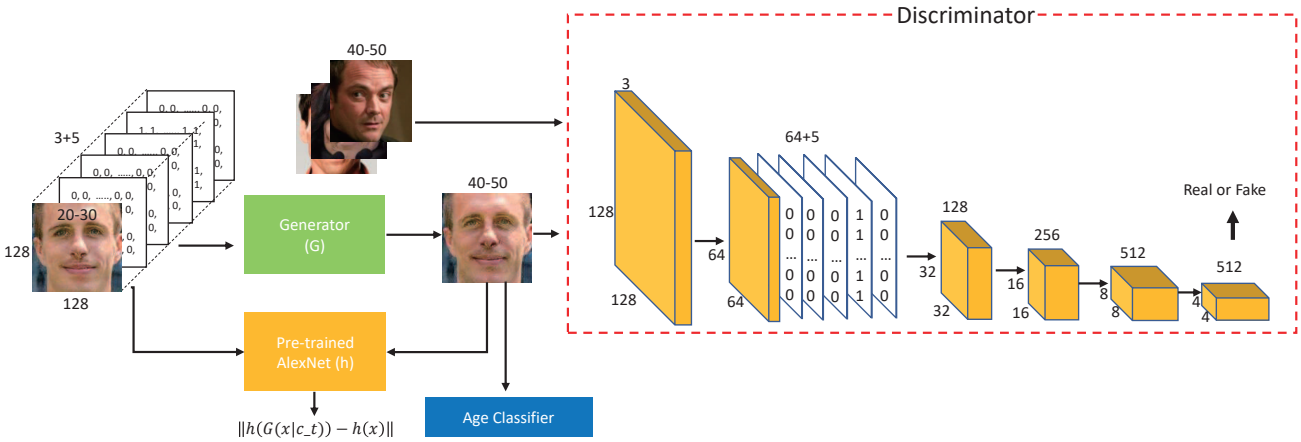

Figure 1. The pipeline of our proposed IPCGANs for face aging. The input image and target age label are concatenated together and then is fed into the generator G. The label is of size $128 \times 128 \times 5$. The discriminator D tries to separate the synthesized faces and faces within the target age group. To preserve the identity information, we enforce the features of synthesized face and input to be similar. We also use an age classifier to force the synthesized face to be with the target age.

To avoid this, in [10], a feed-forward network is adopted. Different from style transfer, which transfers the style of one image to another image, in face aging, it is desirable to transfer the age pattern in the target age group to one face. Therefore, style transfer cannot be directly applied to face aging.

3. Identity-Preserved Conditional Generative Adversarial Networks

3.1. Overview

We divide faces with different ages into 5 nonoverlapping groups. The faces within these 5 groups corresponding to aged 11-20, 21-30, 31-40, 41-50, and 50+, respectively. Given a face image x, we use a code $C_s \in \mathbb{R}^{h \times w \times 5}$ to indicate the age group that x belongs to, in which, h and w represent the height and width of a feature map, 5 is the age group number. Like one-hot code, only one feature map is filled with ones while the rest feature maps are all filled with zeros. Face aging aims to generate a synthesized face $\tilde{x}$ that lies in target age group C_t. It is desirable that the generated face $\tilde{x}$ has the following characteristics: i) $\tilde{x}$ looks realistic; ii) $\tilde{x}$ has the same identity as x; iii) the age of $\tilde{x}$ lies in the target age group C_t. To reach these goals, we propose an Identity-Preserved Conditional GANs (IPCGANs) framework. In our implementation, we train multiple models based on the age group that x belongs to. In other words, the model is only related to the C_s. The model corresponds to C_s can map any face in group C_s to any target age group C_t. For simplicity, We slightly abuse notions and do not specify which C_s age group the model corresponding to in the following sections.

3.2. Identity-Preserved Conditional Generative Adversarial Networks

The whole framework of our Identity-Preserved Conditional Generative Adversarial Networks (IPCGANs) is illustrated in Figure 1. It contains three modules: i) A CGANs module which generates a synthesized face with target age C_t and guarantees $\tilde{x}$ looks realistic; ii) An Identity-Preserved module which guarantees $\tilde{x}$ has the same identity with x; iii) An age classifier module which further enforces $\tilde{x}$ with the desired age C_t.

CGANs based face generation module. Since face aging aims at generating a synthesized face with a target age, we adopt Conditional GANs (CGANs) for face generation. Specifically, we denote y as real faces within the target age group, and denote the distribution of x and y as $p_x(x)$ and $p_y(y)$, respectively. With CGANs, the synthesized face with the target age C_t should not be classified as a faked sample by discriminator D. For real faces, the probability that they belong to real face $D(x|C_t)$ should be high. Besides, D is also responsible for aligning the input label C_t with the generated images. Consequently, we arrive at the following objective function:

$$\min_G \max_D \quad \mathbb{E}_{x \sim p_x(x)}[\log D(x|C_t)] + \mathbb{E}_{y \sim p_y(y)}[\log(1 - D(G(y|C_t))] \tag{2}$$

Similar to the standard GANs [7], the optimization of CGANs in Equation (2) also suffers from instability. Consequently, the generated images are unrealistic and of bad quality. In [12], a Least Squares Generative Adversarial Networks (LSGANs) model is proposed. As shown in [12],

the objective of standard GANs can easily get stacked into a very small loss for the faked samples because the discriminator can easily tell the generated faces and real faces apart. By contrast, LSGANs tries to push both the generated faces and real faces close to the decision boundary and make them indistinguishable. Thus LSGANs is shown to be able to generate high quality images and training is more stable. Therefore, we choose a Conditional LSGANs for our face generation task, which is a special CGANs. Mathematically, the Conditional LSGANs can be formulated as follows:

$$\begin{aligned} L_D =& \frac{1}{2}\mathbb{E}_{x\sim p_x(x)}[(D(x|C_t)-1)^2] \\ &+\frac{1}{2}\mathbb{E}_{y\sim p_y(y)}[(D(G(y|C_t))^2] \\ L_G =& \frac{1}{2}\mathbb{E}_{y\sim p_y(y)}[(D(G(y)|C_t)-1)^2] \end{aligned} \tag{3}$$

To optimize Conditional LSGANs, we use the matching-aware discriminator proposed in [19] which is shown effective for aligning conditions with the generated images.

Identity-preserved module. It is important to preserve the identity information for the synthesized faces. However, the adversarial loss only makes the generator generate samples that follow the target data distribution, consequently, the generated samples can be like any person in the target age group. In other words, adversarial loss alone can not guarantee that the generated samples can preserve the identity information. To keep the identity information for the generated faces, we introduce the following perceptual loss into our face aging objective:

$$L_{identity} = \sum_{x\in p_x(x)} \|h(x) - h(G(x|C_t)\|^2 \tag{4}$$

Here $h(\cdot)$ corresponds to features extracted by a specific feature layer in a pre-trained neural network. The reason of not using mean square error (MSE) between x and its aged face $G(x|C_t)$ in pixel space is that the aged face contains changes in terms of hair color, beard, wrinkles, receding hairline, etc., therefore it is different from x any more. An MSE loss will force $G(x|C_t)$ to be the same as x. However, a perceptual loss encourages the generated images to be close to the features of input face in the same feature space.

Choosing features extracted from a proper layer $h(\cdot)$ is of great importance for preserving the identity information. Experiments in style transfer [10, 6] indicate that lower feature layers are good at keeping the content, while higher layers help keep style related things like color, texture, etc. Even though aged face has the change in terms of hair color, wrinkles, *etc.*, the identity information should not change. Based on this, here we argue that the face content itself represents the identity information, lower feature layer of a pre-trained neural network should be adopted as $h(\cdot)$. To balance the quality of aged images and the identity information of the faces, in Sec. 4 we line search from $fc7$ to $conv2$ of Alexnet pre-trained on ImageNet and empirically set $h(x)$ as the features of $conv5$ layer. Qualitative and quantitative results show that this setting can preserve the identity well and generate diverse aged faces.

Age classification module. To further guarantee the generated faces fall into the target age group C_t, we pre-train an age classifier and use it to identify which age group the face comes from. During the training of our IPCGANs, we fix the parameters of this age classifier and use it to classify the generated face, $G(x|C_t)$. If the generated face is indeed in group C_t, our age classifier gives a small penalty. On the contrary, if $G(x|C_t)$ is not in group C_t, the age classifier will give a big penalty. Here we introduce an age classification loss L_{age} into the objective of IPCGANs. We use L_{age} to represent the age classification loss. We define L_{age} as follows:

$$L_{age} = \sum_{x\in p_x(x)} \ell(G(x|C_t), C_t) \tag{5}$$

Here $\ell(\cdot)$ corresponds to a softmax loss. Through back-propagation, age classification loss forces the parameters of generator to change and generate faces that lie in the correct age group.

Objective function Overall, to generate a face with the target age and the same identity with CGANs, we arrive at the following objective function:

$$\begin{aligned} G_{loss} &= \lambda_1 L_G + \lambda_2 L_{identity} + \lambda_3 L_{age} \\ D_{loss} &= L_D \end{aligned} \tag{6}$$

where λ_1 controls to want extent the input image is aged. λ_2 and λ_3 controls to what extent we want to keep the identity information and let the generated samples fall into the right age group, respectively. In Sec. 4 we empirically find the optimal λ_1, λ_2 and λ_3.

3.3. Network Architecture

The generator and discriminator networks Inspired by the impressive results of style transfer [10] and unpaired image-to-image translation [32], our generator is the same with [32] except the first convolution layer. Our generator receives $128 \times 128 \times 3$ images and $128 \times 128 \times 5$ condition feature maps as input, so we adopt 6 residual blocks in our generator. Like one-hot code, only one feature map is filled with ones while the rest feature maps are all filled with zeros. We inject the conditions before the first convolution

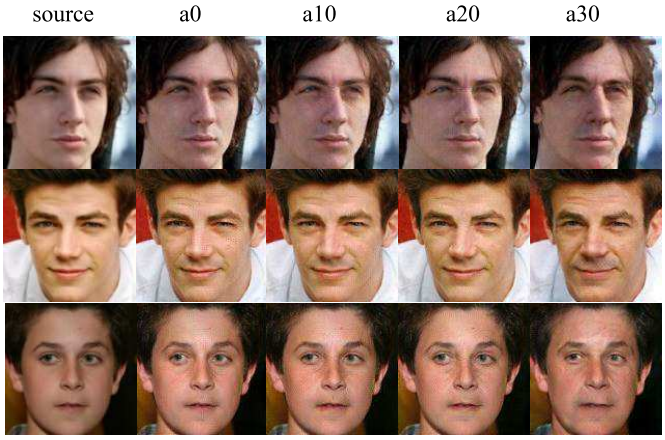

Figure 2. The aging effect of different age classification loss weights. We set $\lambda_1 = 75, \lambda_2 = 5e-5$, input age lies in 11-20 group, target age lies is in 50+ group, throughout. Here we use $conv4$ as the feature layer. $a\mathbf{x}$ means $\lambda_3 = x$. As we can see that as λ_3 grows, the aging effect gets more and more evident. But this trend is limited by the age of the aged images. I.e., if the target age lies in 30-40 group, as λ_3 grows, the aging effect gets more obvious but the aged images will not look like images of group 40-50 or 50+.

layer. The input images and condition feature maps are concatenated together and then the combined feature maps are sent to the first convolution layer.

The architecture of discriminator is adapted from invertible conditional GANs [17] and [32]. The way we inject conditions into the discriminator is exactly the same with [17], which is shown to be able to generate high quality images that are consistent with the conditions. Specifically, as [32] did, we follow the naming convention used in [10]. Let $Conv_k$ represents a 4×4 convolution-Batchnorm-leakyRelu layer with stride 2 and k output channels. The architecture of discriminator is $Conv_{64} - Conv_{128} - Conv_{256} - Conv_{512} - Conv_{512}$. As [18] and [17] suggested, we do not apply Batchnorm on the first $Conv_{64}$ layer and we inject the conditions after this layer. LeakyRelu is with slope 0.2. As the feature map after the $Conv_{64}$ layer is of size 64×64, the size of the condition feature maps that fed into the generator changes from 128 to 64, correspondingly. The $Conv_{64}$ feature maps and conditions are stacked together. The combined feature maps are sent to the $Conv_{128}$ layer.

Age classification network Our age classifier is adapted from Alexnet. The age classifier shares the same architecture from $conv1$ to $pool5$. After that, we add two fully connected layers and a softmax layer, dropout is used to prevent from overfitting.

4. Experiments and Evaluation

In this section, we first introduce the training dataset and image pre-processing details. Then we will evaluate our proposed IPCGANs both qualitatively and quantitatively.

4.1. Dataset

Following the work [28][31], we choose the Cross-Age Celebrity Dataset (CACD) [3] for training and evaluation. CACD contains more than 160,000 face images of 2,000 celebrities with age ranging from 16 to 62. All the images are annotated with age, though not accurate. There are large variations in pose, illumination, expression and even style in this dataset. After face detection, aligning and center cropping, we get 163,104 CACD images whose resolution is 400×400 pixels. We split CACD into two parts, 90% for training and the rest for test. The number of training images of group 11-20, 21-30, 31-40, 41-50, and 50+ is 8,656, 36,662, 38,736, 35,768 and 26,972, respectively.

4.2. Implementation details

We compare our method with the following latest work: age conditional Generative Adversarial Networks (acGANs) [2] and Conditional Adversarial Autoencoder Network (CAAE) [31] which achieve state-of-the-art performance for face aging. All of these methods are based on conditional GANs and are closely related to our method. For acGANs we change the number of age groups from 6 to 5 and replace FaceNet with the pre-trained face VGG net [16]. We use the Tensorflow implementation of L-BFGS-B algorithm and set the maximal iterations to be 1000. CAAE originally has 10 age groups and uses gender information. For fair comparison, we remove the gender information and use 5 age groups instead.

The age classifier is finetuned based on Alexnet on the CACD training set with $200,000$ steps. The learning rate is set as 0.01 at first and is exponentially decreased every 15 epochs. The learning rate decay factor = 0.1, weight de-

cay factor = 0.0005 and batch size = 64. For the training of IPCGANs, we fix the learning rate as 0.001 and use a batch size of 32. The whole training process takes 500,000 steps. As we use BatchNorm throughout, in order to avoid problems with BatchNorm (Running_mean and running_var from BatchNorm layers in test mode might be a bit off from training, which causes significant differences between images generated on training mode and test mode.), we follow the method in [17] to stabilize the BatchNorm layers before using the generator models for image synthesis.

4.3. Qualitative comparison

Following previous work [2] [31] [28], we first qualitatively compare the synthesized faces of different methods. We randomly choose 6 persons from the 11-20 CACD test age group. Figure 3 shows the aged faces of different methods. Since the source code of acGAN is not available, we try our best to implement the original work and tune the parameters to improve the performance. We think we reproduce the same image quality as presented in the original paper of acGANs. For CAAE, we retrain a model with their released code. We can see that images generated by acGANs have lots of artifacts. Besides, acGANs has the danger of losing identities when the target age grows. By contrast, images generated by CAAE look blurry and unrealistic. Due to the use of pixel loss between the input image and its aged ones, the aging effect is not evident. Compared with CAAE and acGANS, the synthesized images of IPCGANs have fewer artifacts, higher image quality and lower possibilities of losing identities.

4.3.1 The effect of identity-preserved module

Figure 4 shows the objective with/without identity-preserved module. Without identity-preserved term, although the adversarial loss makes the input face aged, sometimes the generated images have lots of artifacts and have the danger to lose their identities. With $L_{identity}$, the quality of synthesized images is closely related to which layer is chosen as the feature layer. We fix the other factors unchanged and line search the optimal feature layer from $fc7$ to $conv2$. We can see that as feature layer goes from shallow to deep, the aging effect gets more and more evident and artifacts and distortions will appear. To balance between the image quality and the face aging effect, we empirically set $h(x)$ as the features of $conv5$ layer.

4.3.2 The effect of age classifier module

L_{age} pushes the generator to generate samples that lie in the target age group. Figure 2 shows the effect of age classification term. Keep the other factors unchanged, as the age loss weight λ_3 grows, the aging effect gets more evident.

Table 1. The performance of different methods.

	CAAE	acGANs	IPCGANs
Face verification (%)	91.53	85.83	**96.90**
Image quality (%)	68.85	39.67	**71.74**
Age classification (%)	24.84	**32.70**	31.74
VGG-face score	19.53±1.76	23.42±1.82	**36.33±1.85**
Time cost (s)	0.71	38.68	**0.28**

4.4. Quantitative comparison

4.4.1 User study evaluation

Most existing work quantitatively evaluate the performance of different methods based on user study [31][28]. Following these work, we also conduct user study experiments to compare the quality of faces generated by different methods. Specifically, we invite 80 volunteers who have no knowledge about our work to rate the faces generated by different methods. Different from previous work, besides the image quality evaluation [31][28], we also ask users to conduct face verification task and age classification task for synthesized faces. We randomly select 100 images in the 11-20 age group. Then for each image, we generate 4 aged faces with different target age conditions. Finally, we get 400 aged faces for each aging method. Throughout this part, we use these images for user study evaluation. So the same input $\rightarrow$ output mappings are generated for each model and images presented to all volunteers are the same, which guarantees the fairness of comparison.

Image quality. In this part, we ask volunteers to rate the quality of each face (good or bad). Then we calculate the percentage of images rated as good.

Age classification. Given a synthesized face, we ask volunteers to vote which age group that this face belongs to. By repeating this process for each method, finally, the percentage of faces whose target age agrees with that of user estimation is reported.

Face verification. For each input image, we generate 4 aged images given different age labels. We denote the 4 aged images as age1-4. We form 3 pairs here. (input, age1), (age2, age3), (age4, one randomly selected generated image of other persons). The first 2 pairs are to verify whether the generated images are the same person as the input. The last pair is to verify whether the generated image seems like the other person. Then we ask the users to do face verification task and report the accuracy of different methods. Here the $accuracy = (t_p + t_n)/(N_p + N_n)$. If the trained model is not identity preserved or generates the same person given different inputs, face verification score must be low.

4.4.2 Inception score.

Inception score is another metric used for evaluating the quality of generated images [22]. However, the inception

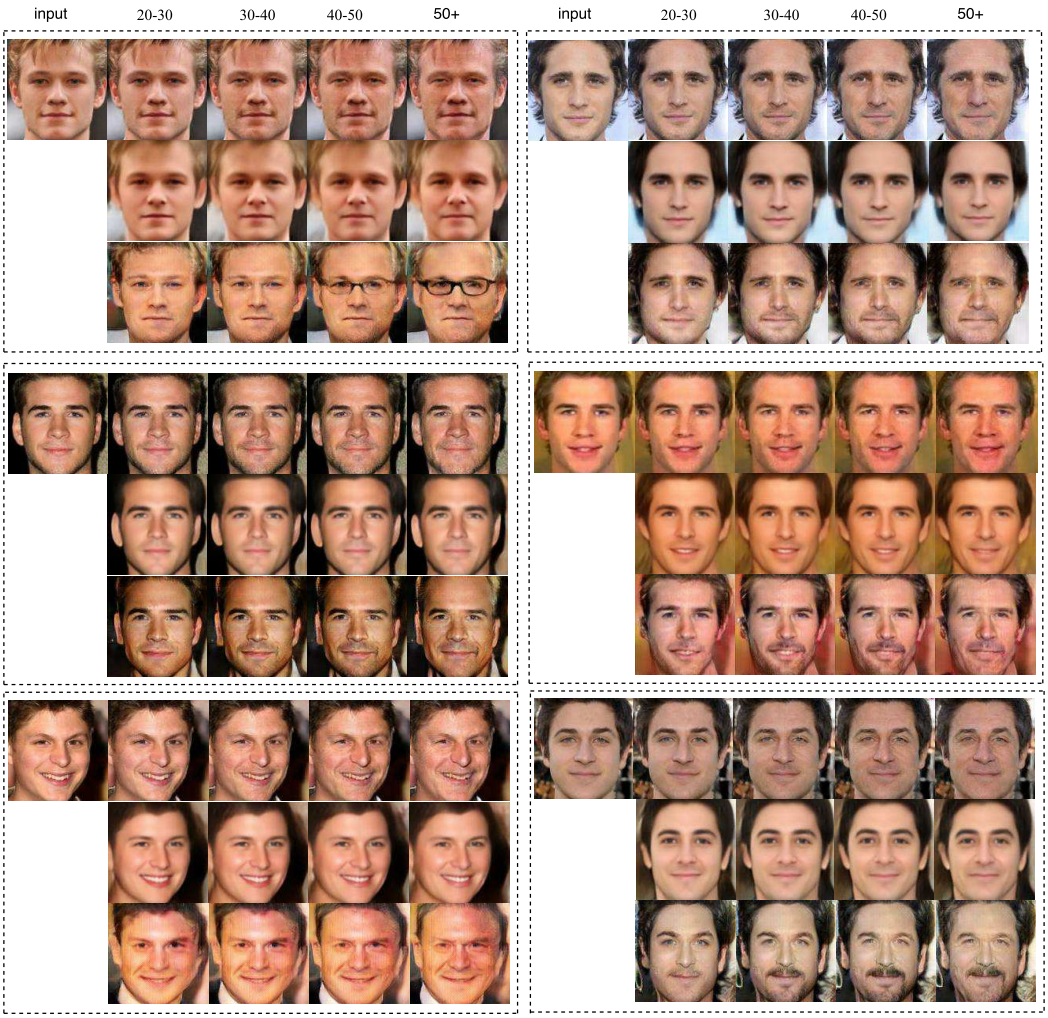

Figure 3. Some synthesized faces generated by different methods. Each dotted box denotes one person's images. In each box, from top to bottom, they are images generated by IPCGANs, CAAE and acGANs. The input age lies in 11-20 age group and target age lies is in 50+ age group. $conv5$ is chosen as the feature layer of $L_{identity}$.

Table 2. The effect of with/without identity-preserved module and age classifier module(%)

age classification		face verification	
with age classifier	w/o age classifier	with identity-preserved term	w/o identity-preserved term
31.37	28.73	**99.07**	98.15

network on ImageNet is trained on 1000 object classes that exclude human faces or human categories. Using the inception score calculated by the network trained on ImageNet to measure the image quality of faces is inappropriate. Instead, we use the pre-trained face VGG net for evaluation. We run the OpenAI source code to compute score. We term this score as VGG-face score.

4.4.3 Computational cost

For fair comparison, here we evaluate the average time cost of generating 4 aged images conditioned on one input image by different methods. We set the maximal iterations of L-BFGS-B to be 1000 and keep the same settings for all methods. Each method is repeated 5 times then the average time is computed.

The performance of all the measurements by different methods is reported in Table 1. We can see that our IPCGANs achieves the highest performance on image quality and face verification. Further, our IPCGAN also achieves the highest VGG-face score, which validates the effectiveness of our method for generating a high quality face with the same identity and target age. As for the computation efficiency,

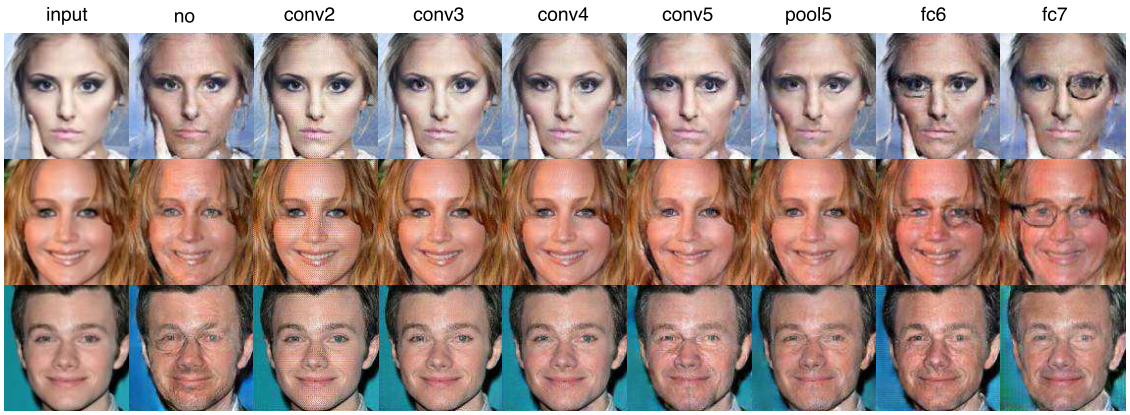

Figure 4. The aging effect with different feature layers. Here we set $\lambda_3 = 0$. The input age lies in 11-20 age group and target age lies is in 50+ age group. 'no' means without the identity-preserving module. As we can see, the lower the layer, the stronger the ability to keep the source image content. As the feature layer goes deeper, the aging effect gets more evident.

our method is nearly $100\times$ faster than acGANs and $2.5\times$ faster than CAAE.

4.4.4 Face recognition with synthesized data

To avoid the suspicion that the limited images used for user study is carefully selected and to further validate the effectiveness of our method, we use the synthesized data to augment the real faces and train a classifier for face recognition. Because of the slow speed of acGANs, We only compare the performance of classifiers trained by using data augmented with synthesized faces generated by CAAE and IPCGANs. For each training image, we generate 4 images given different age conditions. Thus the augmented training set is $4\times$ larger than the original training set. Without data augmentation, the face recognition accuracy with face VGG classifier on the test set is 84.9%. We finetune on this model with the augmented training set, then the best performance achieved by CAAE and IPCGANs is 78.2% and 84.8%, respectively. It shows that after data augmentation, IPCGANs model keeps the face recognition performance while the CAAE model degrades the accuracy greatly.

The performance does not drop after finetuning with 4 times more synthesized data validates that our synthesized images preserve the identity information and the image quality. If synthesized faces cannot keep identities well, so many synthesized faces would reduce the performance (as CAAE does). So face recognition accuracy is another measurement for comparing different face aging models. Meanwhile, this is the first work to use face recognition as a measurement to measure different algorithms. This experiment shows that as a data augmentation method, aged faces generated by our method can not improve the face recognition performance.

4.4.5 The evaluation of identity-preserved module and age classifier module

We also quantitatively evaluate the models with/without identity-preserved term and age classification module by conducting user study based face verification and age classification. The experimental setup is the same with Section 4.4. The results are shown in Table 2. We can see that with age classifier, the age classification accuracy is boosted than that without the age classifier. The identity-preserved module also improves the face verification performance.

5. Conclusion

In light of the success of GANs for image generation, we propose a conditional GANs based face aging approach. The discriminator in CGANs guarantees the consistency between the aged faces and the corresponding target age. To preserve the identity of input images, we force the high level features of input faces and the synthesized faces to be similar. Further, we introduce an age classifier module to force the synthesized faces to be with the target age. In this way, our method can generate high quality faces with the same identity and target age. Both of qualitative and quantitative experiments validate the effectiveness of our approach. Besides, IPCGANs is a general framework. Without any modification, it can be applied to multi-attribute transfer tasks like brown hair to black/blond/gray hair, no beard to beard/5 o'clock shadow/mustache/sideburns, etc. If we remove the condition part, our framework can be used for image translation task, like from RGB domain to near infrared domain.

Acknowledgements. This work is supported by NSFC (NO. 61502304) and Shanghai Subject Chief Scientist (A type) (No. 15XD1502900).

References

[1] G. Antipov, M. Baccouche, and J.-L. Dugelay. Boosting cross-age face verification via generative age normalization. In *International Joint Conference on Biometrics*, 2017.

[2] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. In *IEEE International Conference on Image Processing*, 2017.

[3] B.-C. Chen, C.-S. Chen, and W. H. Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014.

[4] A. Dosovitskiy and T. Brox. Generating images with perceptual similarity metrics based on deep networks. In *Advances in Neural Information Processing Systems*, pages 658–666, 2016.

[5] Y. Fu, G. Guo, and T. S. Huang. Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 32(11):1955–1976, 2010.

[6] L. A. Gatys, A. S. Ecker, and M. Bethge. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*, 2015.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[8] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028*, 2017.

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

[10] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016.

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

[12] 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.

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

[14] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[15] G. Panis and A. Lanitis. An overview of research activities in facial age estimation using the fg-net aging database. In *European Conference on Computer Vision*, pages 737–750. Springer, 2014.

[16] O. M. Parkhi, A. Vedaldi, A. Zisserman, et al. Deep face recognition. In *BMVC*, volume 1, page 6, 2015.

[17] G. Perarnau, J. van de Weijer, B. Raducanu, and J. M. Álvarez. Invertible conditional gans for image editing. *arXiv preprint arXiv:1611.06355*, 2016.

[18] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[19] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text to image synthesis. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 3, 2016.

[20] K. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 341–345. IEEE, 2006.

[21] R. Rothe, R. Timofte, and L. V. Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision (IJCV)*, July 2016.

[22] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, pages 2234–2242, 2016.

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

[24] Z. Shu, E. Yumer, S. Hadap, K. Sunkavalli, E. Shechtman, and D. Samaras. Neural face editing with intrinsic image disentangling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5541–5550, 2017.

[25] J. Suo, S.-C. Zhu, S. Shan, and X. Chen. A compositional and dynamic model for face aging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):385–401, 2010.

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

[27] Y. Tazoe, H. Gohara, A. Maejima, and S. Morishima. Facial aging simulator considering geometry and patch-tiled texture. In *ACM SIGGRAPH 2012 Posters*, page 90. ACM, 2012.

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

[29] W. Wang, Y. Yan, S. Winkler, and N. Sebe. Category specific dictionary learning for attribute specific feature selection. *IEEE Transactions on Image Processing*, 25(3):1465–1478, 2016.

[30] H. Yang, D. Huang, Y. Wang, H. Wang, and Y. Tang. Face aging effect simulation using hidden factor analysis joint sparse representation. *IEEE Transactions on Image Processing*, 25(6):2493–2507, 2016.

[31] Z. Zhang, Y. Song, and H. Qi. Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5810–5818, 2017.

[32] 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.