Triple-GAN: Progressive Face Aging with Triple Translation Loss

Han Fang, Weihong Deng*, Yaoyao Zhong, Jiani Hu
Beijing University of Posts and Telecommunications
{fanghan, whdeng, zhongyaoyao, jnhu}@bupt.edu.cn

Abstract

Face aging is a challenging task which aims at rendering face for input with aging effects and preserving identity information. However, existing methods have split the long term into several independent groups and ignore the correlations of age growth. To better learn the progressive translation of age patterns, we propose a novel Triple Generative Adversarial Networks (Triple-GAN) to simulate face aging. Instead of formulating ages as independent groups, Triple-GAN adopts triple translation loss to model the strong interrelationship of age patterns among different age groups. And to further learn the target aging effect, multiple training pairs are offered to learn the convincing mappings between labels and patterns. The quantitative and qualitative experimental results on CACD, MORPH and CALFW show the superiority of Triple-GAN in identity preservation and age classification.

1. Introduction

Face age synthesis [3] is the process of rendering image with the same identity and changing its age to predict the appearance of any period, which can be also called age progression and regression. Face aging methods have attracted a lot of researchers to synthesize high-quality face with significant age changes and consistent identity in the application of entertainment, social security and cross-age face recognition. For example, it could be helpful to find the missing suspect or victim [5] by generating face in the desired period. However, due to many factors involving living styles, generic plastic surgery, and the shortage of enough labeled data, face age synthesis is a still huge challenge.

Recently, many researchers focus on producing realistic elderly faces. And these methods can be roughly categorized into three parts: physical model-based methods [23], prototype-based methods [9] and deep generative networks [15][30][18][35][28][12][14][33]. The traditional aging methods mostly rely on modeling wrinkles, hair, texture and facial appearance mechanically or use a lot of data to construct prototypes as age patterns. Deep learning methods have recently achieved great success in face aging. By training to learn the specific age patterns and mappings between input faces and target age labels, deep learning methods can generate faces of specific age group directly, which have recently achieved great success in face aging. Although deep learning methods can be easy to learn the age patterns, they can not generate satisfactory results in the desired age group. Besides, the age group based synthesis splits the long-term into several independent groups and add identity preservation between input and output, but ignore the progressive change of age pattern and identity preservation between the synthesized images.

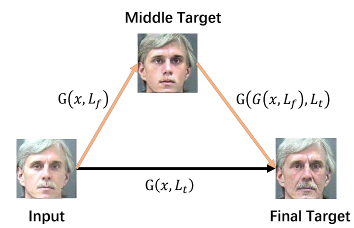


Figure 1. The pipeline of employing triple translation. $G(x, L)$ is the synthesized faces. L_t and L_f represent the different target age groups. Generators with the same parameters are applied to translate synthesized faces of specific age group to another age group. And triple translation loss pushes two faces generated from different paths to keep the same age pattern and identity information.

To solve the mentioned problems, we propose Triple Generative Adversarial Networks (Triple-GAN). We explore to translate different age patterns simultaneously to provide multiple training pairs in adversarial learning. The conditional discriminator can not only discriminate at the real-fake level, but also build efficient mappings between patterns and labels by learning different age patterns jointly. Furthermore, to enhance the performance of generator and model age interrelationship among different age groups, triple translation is added to translate the synthesized face of a specific age to another age. The pipeline is depicted in Figure 1. By employing triple translation loss, synthesized faces in the same target from different paths are forced to be close to each other, so that translation of age patterns can be correlated to better simulate progressive and continuous changes in face aging. Meanwhile, in addition to supervising the input-output distance, a new constraint is introduced into the outputs between different age labels to preserve the identity information between the synthesized faces in different age groups. So aged face cannot lose their identities easily and can be changed in the progressive transfer of age pattern.

Main contributions of our work could be summarized as follows:

1. We propose Triple-GAN which contains triple translation to simulate the real growth of ages. By adopting triple translation loss, the progressive mappings of different age domains are fully correlated. The generator is encouraged to be reusable, generating synthesized faces with the more evident aging effect.

2. Enhanced adversarial loss is adopted to effectively model the complex distribution of age domains. And the identity information between synthesized faces with different target labels is further used to make the translation of age patterns keep progressive and keep the preservation of identity more stable.

3. We conduct identity preservation and age classification for the generated aged faces by the online facial analysis API of Face++ [8] quantitatively. The improvement of age classification accuracy and identity verification confidence on MORPH [20] and CACD [1] have shown the superiority of Triple-GAN. Synthesized images generated from Triple-GAN are also used to augment the real dataset to improve the accuracy of CALFW [34], verifying the effectiveness in cross-age face recognition.

2. Related Work

Recently, face age synthesis has been a more and more popular topic and there are some remarkable progress in the problem of face age synthesis. The published studies on face age synthesis can be roughly categorized into three parts: prototype-based methods, physical model-based methods and deep learning methods.

The prototype-based [9][24] methods use non-parametric model. The faces firstly should be divided into groups according to different ages. The average face of each age group is referred to as prototype and age pattern of specific age group. Shu et al. [21] have proposed an age synthesis method based on dictionaries. Yang et al. [31] have introduced using hidden factor analysis joint sparse representation. These proposed aging methods separately model the stable person dependent properties in a relatively long period and the age-dependent information that gradually changes over time. However, since the age pattern is obtained by the average face, prototype-based methods may ignore the personalized information.

Physical model-based methods pay attention to designing a complex model [23][26][25][22] to imitate the facial appearance and simulate aging mechanisms in terms of hair, muscles and texture for adults and adopting specific transformation on a set of landmarks [2][19] or statistical parameters [11][17] to model age-related shape changes for children. Wu et al. [29] have described a model to simulate expressive wrinkles in 3D animation and skin aging. However, this kind of method needs to construct the parametric model and requires a lot of faces in the same identities under different ages, which is computationally expensive and hard to collect.

Recently, deep learning methods have achieved great success in face age progression and regression. Wang et al. [27] have used a recurrent neural network to make a smooth face age synthesis. Zhang et al. [33] have proposed Conditional Adversarial Auto-encoder (CAAE) to synthesize target age faces with target age labels. Li et al. [13] have introduced the spatial attention mechanism to limit image modifications to regions closely related to age changes. Yang et al. [30] have proposed a multi-pathway discriminator to ensure the synthesized faces present desired aging effects and keep personalized properties. Palsson et al. [18] have presented the CycleGAN based model to further refine the aging effects and use cycle consistency loss to preserve identity information. To make use of the global and local information simultaneously, Li et al. [12] have put forward a novel generator to combine both information to learn the whole facial structure and imitate subtle changes of crucial facial sub-regions. Wang et al. [28] have proposed to impose an identity-preserved term and an age classification term into the objective of GANs.

3. Methodology

3.1. Overview

Due to the limited age information during training, target age pattern cannot be effectively learned and imposed into the generator, resulting in undiscriminating boundaries between different age groups. To tackle this problem, we introduce Triple-GAN, which can learn the effective aging process. As illustrated in Figure 2, our framework contains four components including the generator network G, the pre-trained identity-preserved network, the pre-trained age classification network and the discriminative network D. The datasets are labeled with several age labels L according to 5 age groups: $11-20$, $21-30$, $31-40$, $41-50$, $50+$. The label is one-hot like matrix, which is filled with one in one dimension and zeros in other dimensions to indicate the target age. Given a face image x and concatenate

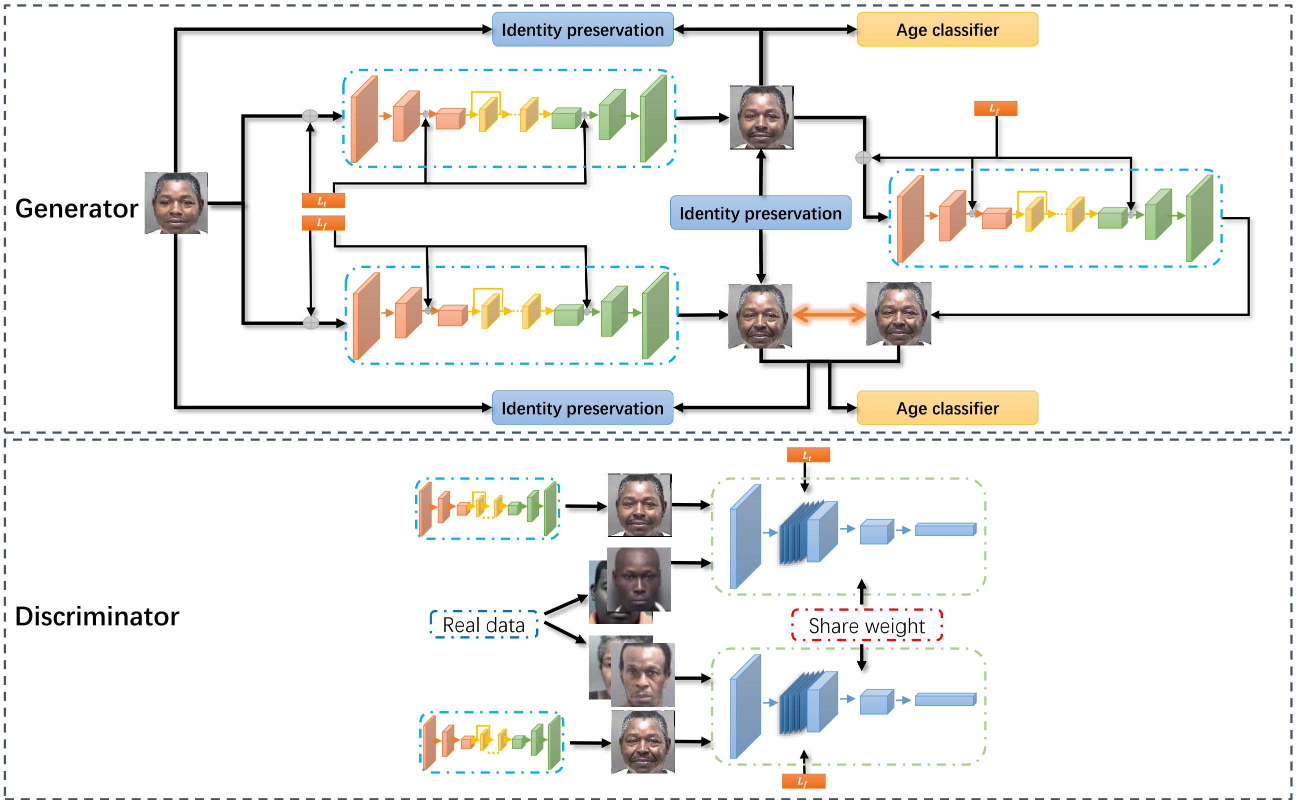


Figure 2. Framework of proposed Triple-GAN for face aging.

with two different target age labels L_t and L_f respectively. Then the combined feature maps can go through several strided convolutional layers to encode into high-level feature space and decode with multi fractionally-strided convolutional layers to get two synthesized faces, which denote as $G(x, L_t)$ and $G(x, L_f)$. By offering combinations of generated faces in different ages to train simultaneously, the discrimination of discriminator can be enhanced. To correlate the two paths, triple translation loss is incorporated to help the generator to focus more on the progressive change in age pattern. To better simulate the realistic face images, the pre-trained age classification network is adopted to push the ages of generated faces to lie to the target group and the identity-preserved network is employed to preserve identity information not only between input and output but also between outputs.

3.2. Network Architecture

Generator G. Inspired by the impressive results of unpaired image-to-image translation, our generator network mainly follows the architecture proposed by Zhu et al. [36], which includes three parts: encoder G_1, residual architecture R and decoder G_2. $128 \times 128 \times 3$ faces and $128 \times 128 \times 5$ label maps are concatenated and fed into G_1 as input and pass through three convolutional layers. The kernel size of first convolutional layer is 7×7 and the last two convolutional layers are filled with 3×3 kernels. Each convolutional layer is followed by one batch-normalization layer and one ReLU layer, which are denoted as the basic block. After passing the second block, the $64 \times 64 \times 5$ label map indicating the same target age group is superimposed to further enhance the effect of age attribute on the generated faces. The final output of G_1 is $G_1(x, L_{128}, L_{64}) \in R^{32\times32\times128}$, which includes image and age information in high-level space. Then 8 residual blocks are adopted to deepen the network. After the residual connections, we use two fractionally-strided convolutional layers to increase the size of feature maps and decode feature maps. Each transposed convolutional layer is also followed by one batch-normalization layer and one ReLU layer. And the same label map is injected into network after the first transposed convolutional block. And the last convolutional layer is filled with 7×7 kernel size followed by Tanh layer to obtain the generated faces.

Discriminator D. The architecture of discriminator is adapted from [28]. The image first passes through the convolutional layer and LeakyReLU layer. Then the label maps of $64 \times 64 \times 5$ are injected to make the network to be able to discriminate whether the images are consistent with the conditions. The convolutional layers followed with the batch normalization layer and LeakyReLU layer form the basic block. All the LeakyReLU are leaky with slop 0.2. And after three blocks and a convolutional layer, the output can be obtained.

3.3. Loss Function

Enhanced Adversarial Loss As depicted in Figure 2, we put forward a new structure of GAN and hope to generate face images of different ages instead of one age. Adversarial loss in most face aging methods based on conditional generated adversarial networks can be expressed as:

$$\min_{G} \max_{D} \; \boldsymbol{E}_y[(D(y_t|L_t)-1)^2] + \boldsymbol{E}_y[D(y_t|L_f)^2] + \boldsymbol{E}_x[D(G(x,L_t)|L_t)^2]. \quad (1)$$

Discriminator tries to learn the correct aging effect by regarding image y_t and corresponding age label L_t as real pair, y_t and age label L_f as fake pair. However, these methods truly learn the aging effect of age label L_t, but ignore the meaning of L_f, which means the networks only know L_f is the fake label but misunderstands what it represents. So one age pattern transfer every batch can make the age boundaries undiscriminating. Since we inject conditions into the discriminator to let discriminator not only work at the real-fake level, but also learn the aging effect. When learning from different transfers of age patterns simultaneously, all the mappings of used age labels and corresponding age information can be effectively extracted. In our Triple-GAN, the generator will produce faces of two ages. All the synthesized faces and real faces of target age group are distinguished in the discriminator simultaneously. Using multiple paths to generate faces of different ages, we offer more training pairs to learn the age mappings. The discriminative age patterns can be learned by adding more boundaries to push them away, which is shown in Figure 3. And we adopt least square loss [16] to push both the generated and real faces close to the decision boundary. The formula of proposed adversarial loss can be formulated as follows:

$$\begin{aligned} L_d =& (D(y_t|L_t)-1)^2 + (D(y_f|L_f)-1)^2 + D(y_t|L_f)^2 \\ &+ D(y_f|L_t)^2 + D(\tilde{x}_{L_t}|L_t)^2 + D(\tilde{x}_{L_f}|L_f)^2, \\ L_g =& (D(\tilde{x}_{L_t}|L_t)-1)^2 + D(\tilde{x}_{L_t}|L_f)^2 \\ &+ (D(\tilde{x}_{L_f}|L_f)-1)^2 + D(\tilde{x}_{L_f}|L_t)^2, \end{aligned} \quad (2)$$

where $G(x, L_t)$ is denoted as $\tilde{x}_{L_t}$.

Age Classification Loss To ensure that the age of synthesized image belongs to the target age group, age classification loss based on pre-trained age classification network is employed, which can be expressed as:

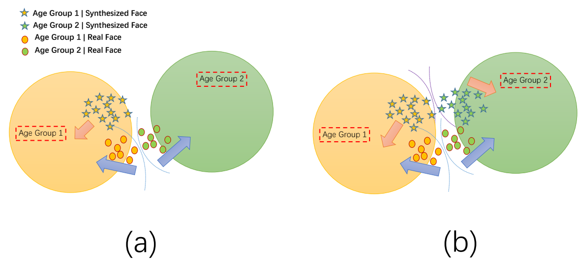


Figure 3. (a) Traditional adversarial loss. (b) Enhanced adversarial loss. Compared with traditional adversarial loss, we offer multiple training pairs to help discriminator to focus on the specific age domain and provide more learned boundaries to push synthesized faces close to the real faces of target age group.

$$L_{age} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}(1\{L=j\} \times \log P), \quad (3)$$

where P is the softmax output which represents the probability and L denotes the target age class. The number of age groups K is 5 in our method.

Identity Preserving Loss In the process of face aging, it is desirable to ensure that the output faces preserve the same identity information as the input face. Adversarial loss will only make the data distribution generated by the generator similar to the target data distribution, but the identity information can not be well preserved. So, to keep the identity information of the generated faces, perceptual loss is adopted to maintain the identity between input and output face images by reducing the differences in the high-dimensional features. The formula can be formulated as:

$$L_{identity}^{input-output} = \sum_{x \in P_x(x)} \|\boldsymbol{I}(x) - \boldsymbol{I}(G(x,L))\|_2^2. \quad (4)$$

Here $\boldsymbol{I}(\cdot)$ indicates the features extracted by a specific feature layer in the pre-trained model. Due to the special pipeline of Triple-GAN, we not only consider minimizing the distance between input and output in a high dimensional space, but also add the constraint to the output between different age labels, which can preserve the identity information between the generated faces in different age groups. The higher similarity between outputs of different ages will make the transfer of age pattern keep progressive and continuous which is hard to lost the identity. So we define the extra identity preserving loss as follows:

$$L_{identity}^{output-output} = \sum_{x \in P_x(x)} \|\boldsymbol{I}((G(x,L_t)) - \boldsymbol{I}(G(x,L_f))\|_2^2. \quad (5)$$

Triple Translation Loss Network can learn the more discriminative age pattern when providing two age patterns

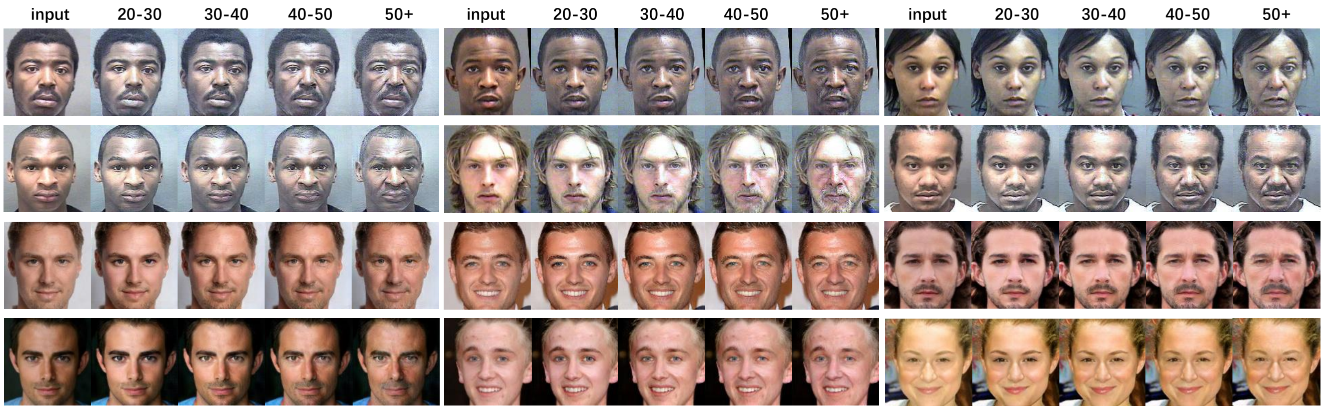


Figure 4. Age-progressed results of Triple-GAN on CACD and MORPH for 6 identities respectively. The leftmost face in each group is the input face and the rest 4 faces are the age-progressed visualizations from young to old in the [21-30], [31-40], [41-50] and 50+ age groups.

simultaneously, but the correlations between two generated images are ignored. In each iteration, the input image of the same identity will produce the output of different ages and the aging process is smoothed by adding identity preservation between different synthesized faces. If the performance of the generator is great enough to generate the face $G(x, L_t)$ and $G(x, L_f)$, $G(x, L_t)$ can be realistic enough to re-sent generator to synthesize $G(G(x, L_t), L_f)$. The faces produced by the generated faces should be completely the same as $G(x, L_f)$. The age-group based methods explicitly split the ages into several groups and ignore the relationship of step-wise age translation. The implicit correlations among all the age domains reveal the progressive change of wrinkles, hair, texture and facial appearance. Unlike previous work [28], which translates faces in five independent domains and only focuses on start and target, triple translation loss focuses more on the mappings among all the domains. The synthesized faces should keep in high visual quality which can be denoted as the input for the second translation. With triple translation loss, the intermediate step can be supervised to follow the process of progressive face aging. So, in order to improve the details of age-progressed face, we add the third generator with the same parameters to fully use the information and make a tight connection between two transfers of age patterns, making up the triple pairs to train the network. Given the input image x, and target labels L_t, the output of the generator is $G(x, L_t)$. Then the output can be re-sent to another location L_f, which should be similar to the result of direct translation to L_f. The formula of triple translation loss can be defined as:

$$L_{triple} = \|G(x, L_t) - G(G(x, L_f), L_t)\|_2^2. \tag{6}$$

By adding supervision, triple architecture forces the faces of the same age groups generated from different paths to retain identity and simulates progressive age changes at different stages. $G(G(x, L_t), L_f)$ is enforced to be totally the same as $G(x, L_f)$, which not only requires the input generated image to maintain great identity characteristics, but also close to the real faces.

3.4. Overall Objective

Triple-GAN will generate three kinds of face images: $G(x, L_t)$, $G(x, L_f)$ and $G(G(x, L_t), L_f)$ and all the synthesized faces are used in identity preservation and age classification. Then our full objective is:

$$\begin{aligned} L_G =& \alpha L_g + \beta L_{identity} + \gamma L_{age} + \lambda L_{triple}, \\ L_D =& \alpha L_d, \end{aligned} \tag{7}$$

where α, β, γ and λ control the weight of four objectives. And $L_{identity}$ includes $L_{identity}^{output-output}$ and $L_{identity}^{input-output}$.

4. Experiment Results

4.1. Data Collection

The two datasets for training GANs are CACD [1] and MORPH [20]. To demonstrate the superiority of proposed method on the cross-age face recognition, CALFW [34] is adopted for the test. Cross-Age Celebrity Dataset (CACD) [1] contains 163,446 faces of 2,000 subjects with age ranging from 16 to 62. Chen et al. [1] selected names from online celebrity database, IMDb.com, and then collected the faces via Google Image Search. MORPH (Album2) [20] is the public available longitudinal dataset which contains 52,099 color faces. The ages of celebrity ranges from 16 to 77 years old. And CALFW [34] contains 3,000 positive pairs and 3,000 negative pairs to enlarge the age gaps, verified in the same way as LFW [7].

4.2. Implementation Details

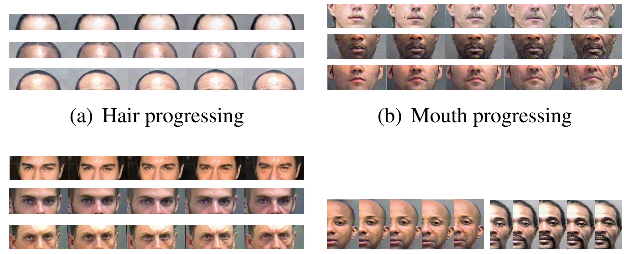
We follow the time period of 10 years for each group as reported in Wang's work [28]. The aim of our work is to synthesize a sequence of age-progressed faces where the ages of input faces are below 20 years old. We use MTCNN [32] to detect landmarks, and align and crop all the face images into $400 \times 400 \times 3$. CACD and MORPH are split into two parts respectively, which contains 80% for training and 20% for the test. Before feeding the images into the networks, the face images are cropped into $227 \times 227 \times 3$ to use in identity preservation term and age classification term, and cropped into $128 \times 128 \times 3$, concatenating the target age group to form feature map of $128 \times 128 \times 8$ as the input of generator. The pre-trained age classifier and identity preservation model are the publicly available AlexNet [10] which we borrow for use from [28]. And we have implemented the following latest work to compare with Triple-GAN: IPCGAN [28] and Conditional Adversarial Auto encoder Network (CAAE) [33] which achieve state of the art on face aging. Since the model of CAAE has divided age into 10 groups and used gender information, we follow Wang et al.'s work [28]. So we use the same 5 groups as our method and remove the gender information for fair comparison. The hyper-parameters of IPCGAN and CAAE are the same as paper reported. To show the effectiveness of triple translation loss, we also report results with two generation pipeline and without third translation, which we called Tuple-GAN. In the training of Triple-GAN, the learning rate is fixed as 0.001, batch size is 16. α, β, γ and λ are set to 75, 0.00005, 35 and 25 in CACD and 75, 0.00005, 35 and 10 in MORPH. For Tuple-GAN, λ is set to 0 and others keep the same. The training process takes 400,000 steps. Discriminator and generator are alternatively updated in every iteration. And the network is first pre-trained in one path for quickly converge and then trained by using overall structure.

4.3. Aging Pattern Simulation

Triple-GAN can synthesize faces of different age groups controlled by the age labels. Although age progression and regression can both be achieved, we focus more on the age progression and show the results on CACD and MORPH respectively. The input faces are under 20 years old. As shown in Figure 4, with age increasing, the evident aging effect and well preserved identities show the aging translations with high visual fidelity. The details of age-progressed results are depicted in Figure 5. Figure 5(a) shows the higher and sparse hairline with age increasing. In Figure 5(b), the beards begin to grow and turn white from black, which evidently demonstrates the change of age pattern. Figure 5(c) indicates that eye line and wrinkle become deeper and more obvious. And in Figure 5(d), the half faces are demonstrated to show the performance globally.

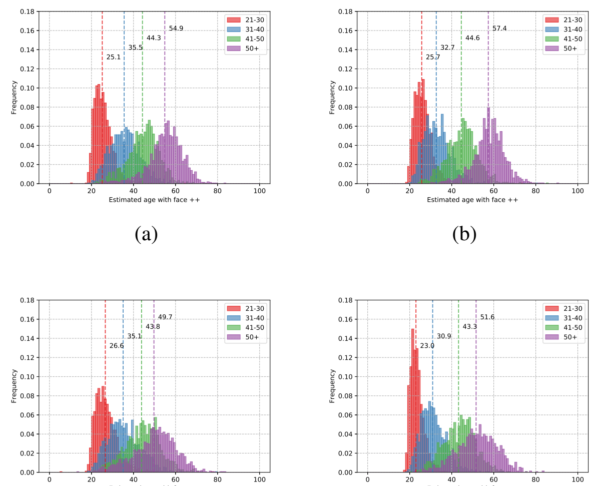
4.4. Aging Classification Accuracy

For a given test dataset which is less than 20 years old, the estimated age of the generated images needs to belong to the target age group. So age classification accuracy is employed to measure the performance. We carry out the online facial analysis API of Face++ [8] to every synthesized face in CACD and MORPH. Test samples in MORPH and CACD whose ages less than 20 are adopted respectively to generate faces in [21-30], [31-40], [41-50] and 50+ age groups. Tuple-GAN, Triple-GAN and IPCGAN [28] are measured in comparison and we report their estimated ages and age classification accuracy in Table 1. It can be seen that the accuracy of Triple-GAN highly outperforms Tuple-GAN and IPCGAN, especially in the age group of 50+, evidently validating the superiority of our method. Furthermore, in Figure 6, both on CACD and MORPH, the age distributions of Triple-GAN in different clusters have better separation than IPCGAN's. It is proved that our age-progressed method has truly captured the target age distribution and generated faces of specific age with high accuracy and discrimination.



(a) Hair progressing (b) Mouth progressing (c) Eye progressing (d) Half face

Figure 5. Illustration of visual fidelity.



(a) (b) (c) (d)

Figure 6. Comparisons in distributions of the estimated ages obtained by Face++ [8]. (a) Triple-GAN, CACD; (b) Triple-GAN, MORPH; (c) IPCGAN, CACD; (d) IPCGAN, MORPH

	MORPH				CACD			
	20-30	30-40	40-50	50+	20-30	30-40	40-50	50+
IPCGAN	22.98	30.96	43.25	51.59	26.57	35.06	43.82	49.71
	85.9	47.9	53.6	59.5	76.8	51.1	50.6	54.5
Tuple-GAN	25.41	32.19	43.26	55.31	25.68	36.35	46.05	54.26
	87.9	50.5	48.5	80.2	82.1	51.5	51.2	76.1
Triple-GAN	25.72	32.64	44.56	57.41	25.12	35.54	44.26	54.88
	87.1	**54.0**	**56.3**	**87.7**	**85.1**	**55.1**	**57.7**	**79.8**

Table 1. Age estimation and accuracy (%) on MORPH and CACD. The first row of each method shows the estimate ages and the second row presents the age classification accuracy (%).

		MORPH				CACD			
		20-30	30-40	40-50	50+	20-30	30-40	40-50	50+
IPCGAN	Test 20-	95.74	95.10	92.53	90.38	95.30	94.06	92.43	**92.66**
	20-30	-	94.98	91.58	88.60	-	93.68	90.13	89.69
	30-40	-	-	94.02	91.30	-	-	93.80	91.63
	40-50	-	-	-	93.68	-	-	-	93.48
Tuple-GAN	Test 20-	95.74	95.10	**93.71**	90.94	95.40	94.16	92.26	92.03
	20-30	-	95.45	93.74	90.90	-	94.37	91.23	89.72
	30-40	-	-	95.46	92.87	-	-	94.57	**92.82**
	40-50	-	-	-	**94.78**	-	-	-	94.24
Triple-GAN	Test 20-	**95.88**	**95.12**	93.66	**91.12**	**95.40**	**94.38**	**92.96**	92.42
	20-30	-	**95.58**	**94.05**	**91.40**	-	**94.82**	**91.96**	**90.26**
	30-40	-	-	**95.51**	**93.10**	-	-	**94.89**	92.41
	40-50	-	-	-	94.76	-	-	-	**94.33**

Table 2. Face verification confidences (%) on MORPH and CACD.

4.5. Identity Preservation

The other objective is that the identity of individual remains the same in the process of age progression, which will not confuse two different identities. Face++ [8] is used to carry out to view the similarity between two faces. We apply comparisons not only between the input faces and the age-progressed results but also between synthesized faces. The faces in every pair are in some identities and different ages and we adopt the score from Face++ [8] as the identity verification confidence. As depicted in Table 2, the identity verification confidences are decreasing with face aging, which means the age pattern actually changes the appearance. However, our proposed Triple-GAN preserves identity information better than others in nearly all cases.

4.6. Contribution of Triple Translation Loss

Triple architecture requires that the generator to use synthesized face as input to generate faces of different target ages. And L_{triple} minimizes the distance between the synthesized faces in the same target age group which are produced in different paths, making the generated faces more realistic and change of age pattern more progressive. Keep other hyper-parameters unchanged and set λ to 0, we report the results of Tuple-GAN in Table 1 and 2. The age accuracy and identity verification confidence achieved on MORPH and CACD are lower than the results obtained by Triple-GAN. This could be caused that the synthesized wrinkles are less clear and the faces look relatively unnatural. Without the supervision of triple translation loss, the distance of domains between input and output is still so far, resulting in the lost subjects and messy faces. The triple translation can enhance the correlations among age domains and solve the problem of learning age patterns independently.

4.7. Compare with Prior Work

To fully demonstrate the effectiveness of proposed Triple-GAN, we compare our method with several prior works: CONGRE [22], HFA [31], CAAE [33], IPCGAN [28], GLCA-GAN [12] and pyramid-GAN [30] which signify state of the art. Figure 7 demonstrates some samples. Furthermore, the faces of IPCGAN and CAAE are implemented by ourselves as we mentioned before. And for a fair comparison, the same faces have been chosen with CONGRE, HFA, GLCA-GAN and pyramid GAN as our input and we directly cite their synthetic results as the baseline. It can be seen that the traditional face aging methods CON-

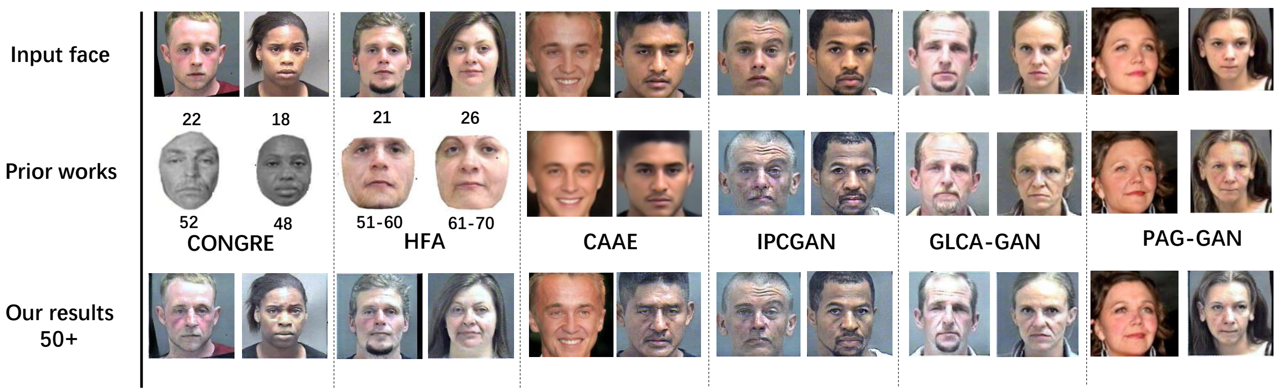
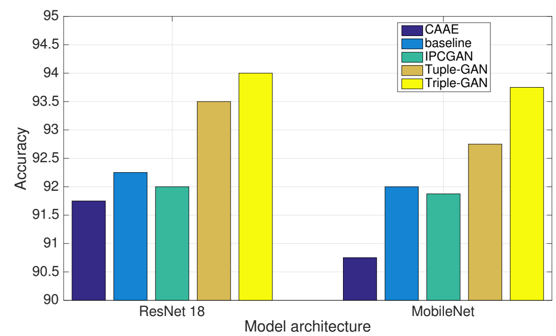


Figure 7. Comparison to prior works including CONGRE [22], HFA [31], CAAE [33], IPCGAN [28], GLCA-GAN [12] and Pyramid-GAN [30]

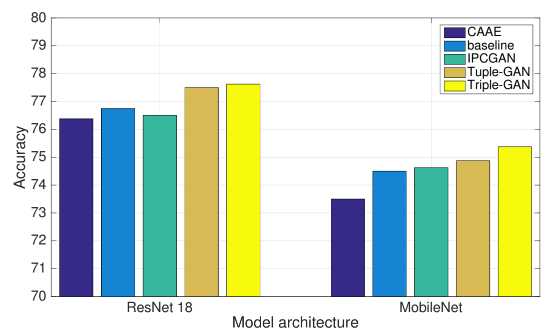
GRE [22] and HFA [31] can only generate subtle aging effects with tight facial area, while the GAN-based algorithm can simulate the change of age pattern on the entire face. Compared with our work, the faces generated by CAAE [33] are not photo-realistic. And other GAN-based methods can simulate the aging effect well. However, because we impose progressive information among age groups, our model is more natural and realistic in some details, such as wrinkle, color and beard.

4.8. Cross-age Face Recognition

To further compare with the global quality of the generated face images quantitatively, cross-age face recognition is employed to indicate the performance. Test data less than 20 years old are used to generate the faces of the target group. And the synthesized data generated by different methods will augment the real data respectively to train the discriminating face classifier. We mainly compare with 5 situations: baseline (which only contains real data), CAAE [33], IPCGAN [28], Tuple-GAN and Triple-GAN. And to show that our method is model agnostic, ResNet-18 [4] and MobileNets [6] are used. Considering the limited number of training set, we only choose 800 pairs of CALFW [34] randomly as the test set, which contains 400 positive pairs and 400 negative pairs. The result is demonstrated in Figure 8, which reveals that Triple-GAN enriches age information while preserves identity information well.



(a)



(b)

Figure 8. Recognition results on subset of CALFW [34] with different methods. Using real dataset for training to get the baseline and using faces generated by different face aging synthesis methods to augment real training set. (a) CACD [1]; (b) MORPH [20].

5. Conclusion

This paper has proposed a novel face aging method called Triple-GAN for synthesizing age-progressed faces. Triple-GAN has adopted enhanced adversarial loss to not only discriminate the realism of synthesized faces, but also learn effective mappings with more age boundaries. Instead of formulating ages as independent groups, triple translation loss has been added to further model the complex correlation of multiple age domains and simulate more realistic age growth, further enhancing the superiority of generator. Several quantitative and qualitative experiments conducted on MORPH, CACD and CALFW have demonstrated the effectiveness of our proposed method.

References

[1] Bor-Chun Chen, Chu-Song Chen, and Winston H Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *European Conference on Computer Vision*, pages 768–783. Springer, 2014.

[2] Douglas DeCarlo, Dimitris Metaxas, and Matthew Stone. An anthropometric face model using variational techniques. In *SIGGRAPH*, volume 98, pages 67–74, 1998.

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

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[5] Catherine M Hill, Christopher J Solomon, and Stuart J Gibson. Aging the human face-a statistically rigorous approach. 2005.

[6] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[7] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. 2008.

[8] Megvii Inc. Face++ research toolkit. `www.faceplusplus.com`.

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

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[11] Andreas Lanitis, Christopher J. Taylor, and Timothy F Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on pattern Analysis and machine Intelligence*, 24(4):442–455, 2002.

[12] Peipei Li, Yibo Hu, Qi Li, Ran He, and Zhenan Sun. Global and local consistent age generative adversarial networks. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 1073–1078. IEEE, 2018.

[13] Qi Li, Yunfan Liu, and Zhenan Sun. Age progression and regression with spatial attention modules. *arXiv preprint arXiv:1903.02133*, 2019.

[14] Si Liu, Yao Sun, Defa Zhu, Renda Bao, Wei Wang, Xiangbo Shu, and Shuicheng Yan. Face aging with contextual generative adversarial nets. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 82–90. ACM, 2017.

[15] Yunfan Liu, Qi Li, and Zhenan Sun. Attribute enhanced face aging with wavelet-based generative adversarial networks. *arXiv preprint arXiv:1809.06647*, 2018.

[16] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2794–2802, 2017.

[17] S Mukaida. Facial image synthesis using age manipulation based on statistical feature extraction. *VIIP2002, Spain*, 2002.

[18] Sveinn Palsson, Eirikur Agustsson, Radu Timofte, and Luc Van Gool. Generative adversarial style transfer networks for face aging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2084–2092, 2018.

[19] Narayanan Ramanathan and Rama Chellappa. Modeling age progression in young faces. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 387–394. IEEE, 2006.

[20] Karl Ricanek and Tamirat Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 341–345. IEEE, 2006.

[21] Xiangbo Shu, Jinhui Tang, Hanjiang Lai, Luoqi Liu, and Shuicheng Yan. Personalized age progression with aging dictionary. In *Proceedings of the IEEE international conference on computer vision*, pages 3970–3978, 2015.

[22] Jinli Suo, Xilin Chen, Shiguang Shan, Wen Gao, and Qionghai Dai. A concatenational graph evolution aging model. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2083–2096, 2012.

[23] Jinli Suo, Song-Chun Zhu, Shiguang Shan, and Xilin Chen. A compositional and dynamic model for face aging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):385–401, 2010.

[24] Bernard Tiddeman, Michael Burt, and David Perrett. Prototyping and transforming facial textures for perception research. *IEEE computer graphics and applications*, 21(5):42–50, 2001.

[25] James T Todd, Leonard S Mark, Robert E Shaw, John B Pittenger, et al. The perception of human growth. *Scientific american*, 242(2):132–144, 1980.

[26] Ming-Han Tsai, Yen-Kai Liao, and I-Chen Lin. Human face aging with guided prediction and detail synthesis. *Multimedia tools and applications*, 72(1):801–824, 2014.

[27] Wei Wang, Zhen Cui, Yan Yan, Jiashi Feng, Shuicheng Yan, Xiangbo Shu, and Nicu Sebe. Recurrent face aging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2378–2386, 2016.

[28] Zongwei Wang, Xu Tang, Weixin Luo, and Shenghua Gao. Face aging with identity-preserved conditional generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7939–7947, 2018.

[29] Yin Wu, Nadia Magnenat Thalmann, and Daniel Thalmann. A plastic-visco-elastic model for wrinkles in facial animation and skin aging. In *Fundamentals of Computer Graphics*, pages 201–213. World Scientific, 1994.

[30] Hongyu Yang, Di Huang, Yunhong Wang, and Anil K Jain. Learning continuous face age progression: A pyramid of gans. *arXiv preprint arXiv:1901.07528*, 2019.

[31] Hongyu Yang, Di Huang, Yunhong Wang, Heng Wang, and Yuanyan Tang. Face aging effect simulation using hidden factor analysis joint sparse representation. *IEEE Transactions on Image Processing*, 25(6):2493–2507, 2016.

[32] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

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

[34] Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. *arXiv preprint arXiv:1708.08197*, 2017.

[35] Siyu Zhou, Weiqiang Zhao, Jiashi Feng, Hanjiang Lai, Yan Pan, Jian Yin, and Shuicheng Yan. Personalized and occupational-aware age progression by generative adversarial networks. *arXiv preprint arXiv:1711.09368*, 2017.

[36] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.