


Article

Bidirectional Face Aging Synthesis Based on Improved Deep Convolutional Generative Adversarial Networks

Xinhua Liu ^{1,2,*} , Yao Zou ^{1,2,*}, Chengjuan Xie ^{1,2}, Hailan Kuang ^{1,2} and Xiaolin Ma ^{1,2}

¹ School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China; xiejc@whut.edu.cn (C.X.); kuanghailan@whut.edu.cn (H.K.); maxiaolin0615@whut.edu.cn (X.M.)

² Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Wuhan University of Technology, Ministry of Education, Wuhan 430070, China

* Correspondence: liuxinhua@whut.edu.cn (X.L.); zy772052352@whut.edu.cn (Y.Z.)

Received: 10 January 2019; Accepted: 15 February 2019; Published: 18 February 2019



Abstract: The use of computers to simulate facial aging or rejuvenation has long been a hot research topic in the field of computer vision, and this technology can be applied in many fields, such as customs security, public places, and business entertainment. With the rapid increase in computing speeds, complex neural network algorithms can be implemented in an acceptable amount of time. In this paper, an optimized face-aging method based on a Deep Convolutional Generative Adversarial Network (DCGAN) is proposed. In this method, an original face image is initially mapped to a personal latent vector by an encoder, and then the personal potential vector is combined with the age condition vector and the gender condition vector through a connector. The output of the connector is the input of the generator. A stable and photo-realistic facial image is then generated by maintaining personalized facial features and changing age conditions. With regard to the objective function, the single adversarial loss of the Generated Adversarial Network (GAN) with the perceptual similarity loss is replaced by the perceptual similarity loss function, which is the weighted sum of adversarial loss, feature space loss, pixel space loss, and age loss. The experimental results show that the proposed method can synthesize an aging face with rich texture and visual reality and outperform similar work.

Keywords: face-aging synthesis; GAN; DCGAN; latent vector; perceptual similarity loss

1. Introduction

With the increase of age, the individual's facial features will change significantly. Compared with other changes, facial appearance changes show some unique features. For example, aging variations are specific to a particular individual; they occur slowly and are significantly affected by other factors, such as health, gender, and lifestyle. In particular, aging is an irreversible and inevitable process [1].

A person's picture with a certain time span in the FG-NET face image database is shown in Figure 1.



Figure 1. A person's picture with a certain time span.

As an important personal characteristic, human age can be directly inferred from different patterns of facial appearance. With the rapid development of computer graphics and machine vision,

computer-based face-aging synthesis has become a particularly popular topic in recent years. The technology of simulating the aging process of a human face has wide application prospects such as safety control, supervision and monitoring, biometrics, entertainment, and cosmetology [2]. However, this technology remains challenging. Firstly, because of the complex structure of the face, a slow and varying aging process, and the variety of reasons that are behind aging, everyone has a unique way of aging. Secondly, the existing facial data sets are confused and inconsistent, such as facial expression, posture, occlusion, and great differences in lighting conditions. Finally, using the existing data sets, it is difficult to meet the requirements of various methods.

In order to maintain more personalization and acquire a clear personalization in an aging face, we propose an improved deep convolution generation antagonism network framework (DCGAN). Firstly, the image is input into the encoder to obtain its personality characteristics. Then, through the connector, personality traits are associated with age and gender traits in order to input more important learning materials for the generator. At the same time, we use the existing mature age estimator to calculate the age loss in the training process to better understand the aging synthesis. In addition, we use the discriminator to distinguish the input image and the generated image to optimize the objective function, which makes it more and more difficult for the discriminator to determine whether the generated image is the image generated by the generator. In addition, in order to generate more detailed and realistic aging facial images, we use a perceptual similarity [3] measure to replace the original single adversarial loss. The specific objective function is the weighted sum of the adversarial loss, feature space loss, pixel space loss, and age loss. The combination of multiple losses makes the final image better.

In general, the contributions of this paper can be summarized as follows: Firstly, we devise special encoders and connectors in order to extract facial personality features, and then, make full use of age and gender information for network training. Secondly, the perceptual similarity measure is applied to the objective function of the proposed model to obtain the aging surface with clearer lineament and facial details. Finally, the proposed method in this paper has strong robustness in posture and occlusion.

2. Related Work

2.1. Face Detection

Face detection is a key link in an automatic face recognition system. It refers to searching any given image with a certain strategy to determine whether it contains a face or not, and if so, returning the position, size, and pose of a face. Suk-Ju Kang et al. proposed a new multi-user eye tracking algorithm based on position estimation [4]. Tsung-Yi Lin et al. designed and trained a simple dense detector called RetinaNet [5]. RetinaNet is able to match the speed of previous one-stage detectors while surpassing the accuracy of all existing state-of-the-art two-stage detectors.

2.2. Face Progression and Regression

Conventional methods of facial-aging synthesis can be roughly classified as the physical model-based method [6] and the prototype-based method [7]. The former takes full account of facial features and geometric construction and establishes parametric models of facial geometric features, muscles, and wrinkles. The latter divides all facial pictures into diverse age groups. Considering the average face of each age group as the prototype, the distinction between the prototypes is treated as the aging pattern. Nevertheless, as a prototype of aging, the normal facial texture is too smooth to capture high-frequency details such as spots and wrinkles. Park, Tong, and Jain ameliorated this technique and made it applicable to a three-dimensional facial-aging field [8]. By building shape space and texture space models, we can get three-dimensional face-aging simulation methods, and some progress has been achieved in this. With the wide application and rapid development of in-depth learning in the image processing domain, a facial-aging method based on a recursive neural network

(RNN) comes up, in which RNN is applied to age mode conversion to much better retain personality characteristics and make full use of the correlation between adjacent age groups [9].

2.3. Generative Adversarial Network

Generative adversarial network (GAN) [10], as a generation model, has attracted widespread attention in academic circles since it was first proposed by Professor Ian Goodfellow of the University of Montreal in 2014. An age-conditional generative adversarial network (Age-cGAN) [11] is proposed by Grigory et al. to preserve recognizable personality features in principle. It is the first time that GAN has been applied to face aging of an appointed age group.

The initial GAN model consists of a generating network and a discriminant network. The network captures the distribution of sample data and generates forged samples. The objective function of the network is a minimax game process to distinguish whether the input is real samples or forged samples. This optimization can be regarded as a minimax two-player game, as in Equation (1), in which z is a vector randomly sampled from a known simple distribution $p_z(z)$, θ_G and θ_D are parameters of generator G and discriminator D .

$$\min_{\theta_G} \max_{\theta_D} V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

Compared with other models, GAN can generate sharper and clearer images without Markov chain or approximate maximum likelihood estimation, avoiding the difficult problem of approximate calculation of probability. However, GAN also has some problems, the most important of which is the collapse of GAN training and the uncontrollable problem of a model which is too free to be controlled. Because the optimization process of GAN is defined as a minimax problem and there is no loss function, it is difficult to distinguish whether the optimization is progressing or not in the process of network training, and the runaway problem may occur in the training process. As a result, the generator will only generate the same sample, the network cannot continue to learn, and the discriminator will lose its function. In addition, GAN does not need to set up a sample distribution model in advance, which is not only a major advantage of GAN but also has some associated problems, for instance, GAN is too free, there is a lack of guidance for data generation, and it sometimes generates fantastic and difficult to understand images.

Over the past three years, researchers have proposed several ways to improve the original GAN from different perspectives. Here are some of the more famous improved GANs in general chronological order. To solve the problem that GAN is too free, Conditional GAN (CGAN) [12] guides the generation of data by adding additional conditional information to GAN, and takes conditional information as the input of discriminator and generator. DCGAN [13] is a very successful network which combines a convolutional neural network with a countermeasure network. Convolution and micro-step convolution neural networks are used to replace discriminant networks and generating networks in GAN, respectively, to enhance the ability of image feature expression. This improved convolution or deconvolution network architecture has basically become the standard structure of GAN design. Afterward, this convolution structure has been used in the improved generation countermeasure network. Wasserstein GAN (WGAN) [14] introduces a new metric distance, the Wasserstein distance, also known as Earth-Mover (EM) distance, which can theoretically solve the problem of GAN gradient disappearance. Combining with WGAN, Boundary Equilibrium Generative Adversarial Network (BEGAN) [15] proposes a new equalization method for balancing generators and discriminators in training and proposes a new approximate metric convergence method to achieve fast and stable training and generate high-quality visual images. A Conditional Adversarial AutoEncoder (CAAE) is proposed by Zhang et al. [16] to generate more realistic facial images by simulating age progression and regression.

3. Proposed Method

3.1. Framework

The framework of the proposed face-aging method in this paper is illustrated in Figure 2. The frame diagram mainly consists of Encoder (E), Connector (C), Generator (G), and Discriminator (D). The function of E is to map the high-dimensional facial image to the personal latent vector and then extract the personality vector z of human faces. C merges z and y to generate a connection vector, where y is composed of age label (a) and gender label (g). The reason for choosing a gender label is to let it guide the face-aging synthesis more purposefully, thereby the result becomes more authentic. The component face image will be generated by inputting z and y into G. After that, D will distinguish the generated faces from input faces in order to generate realistic and reasonable faces.

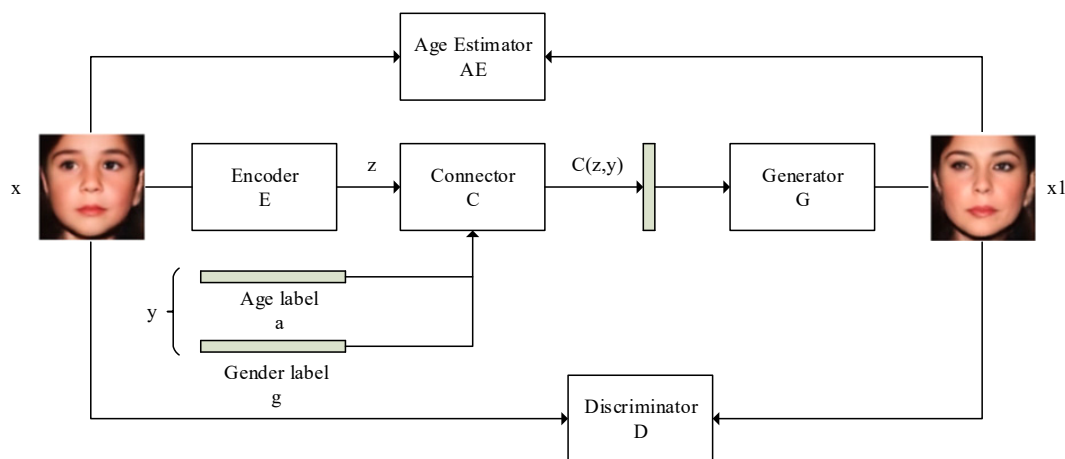


Figure 2. The framework of our method.

The original three-channel image of 128×128 is input into E, and its output is a vector of 1×60 , which contains the personality characteristics of the face. By connecting the 1×60 eigenvector with the replicated age vector and gender vector in C as the input of G, and then G will output a 128×128 three-channel composite image. D will output a probability value for the original image and the composite image, respectively, which represents the probability that the input of the discriminator belongs to a real face image.

3.2. Encoder

When the dimension of the input vector is higher than the dimension of the output vector, the neural network is equivalent to an encoder. In the face-aging synthesis task, for the DCGAN model, the task of the generator is to map a “noise” vector z and a condition vector y to a face image. The function relationship is expressed as:

$$x_1 = G(C(z, y)) = G(E(x), y) \quad (2)$$

where G represents the generator. To ensure that the generated face image is the same person as the input image x , the z vector here must be a representation of the input image x . The purpose of designing the convolutional encoder E is to map the input face image into the hidden space. Feature vector z :

$$z = E(x) \quad (3)$$

The convolutional encoder designed in this paper is a full convolutional network. The whole network replaces the pooling layer with stride convolution, convolves the RGB (Red, Green, Blue)

face image of the 128×128 pixels layer by layer, and finally outputs a low-dimensional feature vector through full connection. The network structure of the convolutional encoder is shown in Figure 3.

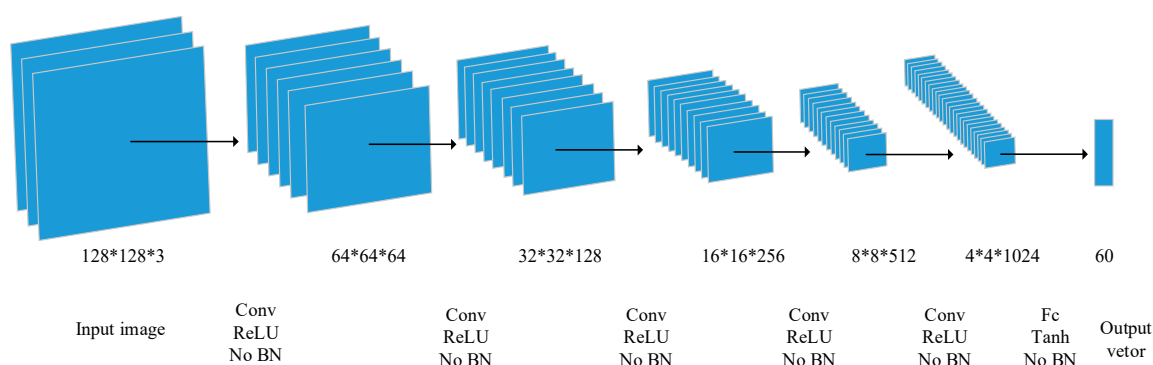


Figure 3. Convolutional network structure of the encoder.

In the convolution layer, the convolution cores of size 5×5 are used, and the steps are set to 2×2 . Batch Normalization (BN) is not used in a convolution coder. Although batch normalization helps to accelerate the convergence speed of the model, there are also problems of sampling oscillation and model instability. From the above table, we can see that in a convolutional coding network, except for the Tanh function used in the last layer of the full connection layer activation function, the ReLU function is used in the other layers.

3.3. Generator

When the dimension of the input vector is lower than the dimension of the output vector, the neural network is equivalent to a decoder. The generator G acts as a decoder corresponding to the convolutional encoder E and takes the face feature vector z , the age vector a , and the gender vector g as input, and the face image is reconstructed from the feature information. The gender condition is added because the aging characteristics of different genders are very different. The aging synthesis of a face is carried out on the basis of clear gender, which can avoid the influence of gender on the aging result. The network structure of the generator is shown in Figure 4.

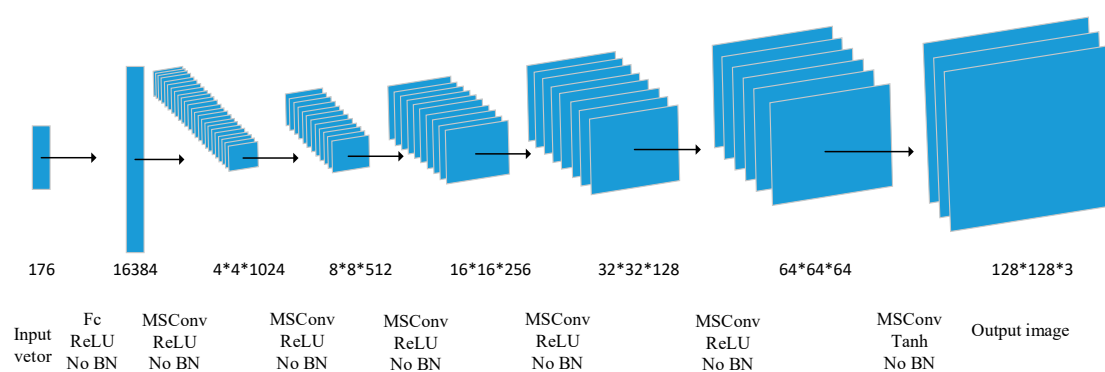


Figure 4. Deconvolution network structure of the generator.

The input to the generator is the merge vector of feature vector z , age vector a , and gender vector g . From the structure diagram of the convolutional encoder, z is a 60-dimensional feature vector, and the age information is an eight-dimensional one-hot vector. The gender information is a two-dimensional one-hot vector. If directly merged, the age condition and gender condition will have little effect on the generator. In order to balance the influence of eigenvectors and conditional vectors on the composite image, the age vector a is copied seven times before merging to obtain a 56-dimensional vector, and the gender vector g is copied 30 times to obtain a 60-dimensional vector.

Then, the conditional input of the generator is a 116-dimensional vector, and the eigenvector z is combined to obtain a 176-dimensional input.

The most important operation of generating a network is the Fractional-Strided Convolution, which is also considered to be deconvolution in many places. In the micro-step convolution operation, it is adopted. A convolution kernel of size 5×5 with a step size of 2×2 . Similar to the convolutional coding network, batch normalization is not used in the generation network, and the Relu activation function is used for all layers except the output layer using the Tanh activation function.

3.4. Discriminator

The role of the discriminator is to distinguish between the real face image and the synthetic face image and, finally, output a scalar value indicating the probability that the discriminator's input image is a real face image. The network structure of the discriminator is shown in Figure 5.

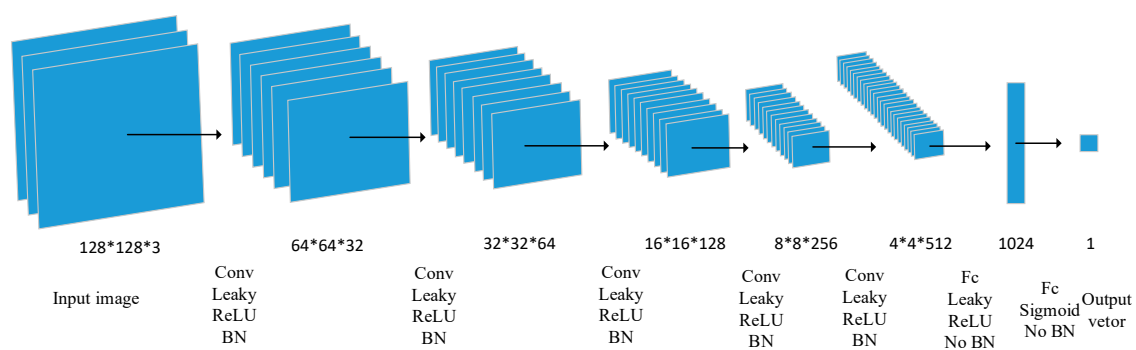


Figure 5. Convolution network structure of the discriminator.

As can be seen from Figure 5, the input is an RGB face image (real image or composite image) of size 128×128 pixels, and the output is a scalar value in the range of (0, 1). The constraint is connected to the first convolutional layer according to the design rules of the condition GAN. Specifically, after the input image passes through the first convolutional layer, a feature map of 16 pixels is output and then connected to the conditional feature map after the extended copy to obtain a feature map of 32 pixels. After the conditional connection is successful, the 32 feature maps are convoluted and fully connected, and finally, a scalar value representing the probability is output.

Unlike the convolutional encoder and generator, the discriminant network uses the Leaky ReLU activation function in addition to the last layer using the Sigmoid activation function. Furthermore, in the discriminant network, the convolutional layer uses batch normalization.

3.5. Loss Function

The objective function of the original GAN is a zero-sum game about the discriminator and the generator, and it is also a maximal-minimization problem. The objective function of the original GAN, is:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (4)$$

where G represents the discriminating network, D represents the generating network, $G(z)$ is the output of the generating network, z is a noise vector, and $D(x)$ indicates the probability that the input is a true data sample. And $p_{data}(x)$ is the data distribution which represents x , $p_z(z)$ depicts the distribution of the noise vector z .

For the aging synthesis task of this paper, the objective function of GAN becomes:

$$\min_G \max_D \{ E_{x,y \sim p_{data}(x,y)} [\log D(x,y)] + E_{x,y \sim p_{data}(x,y)} [\log(1 - D(G(E(x), y)))] \} \quad (5)$$

where x is the input face image and y is the condition constraint (age and gender condition), indicating the face sample distribution. E , D , and G are represented as convolutional encoders, discriminators, and generators, respectively.

In the aging synthesis process, the purpose of the discriminator D is to distinguish the real face image from the synthesized face image under given constraints, so the objective function of the discriminator is:

$$\max_D \{E_{x,y \sim p_{data}(x,y)} [\log D(x,y)] + E_{x,y \sim p_{data}(x,y)} [\log(1 - D(G(E(x),y)))]\} \quad (6)$$

Converting the formula above to a minimized form gives:

$$\min_D \{-\{E_{x,y \sim p_{data}(x,y)} [\log D(x,y)] + E_{x,y \sim p_{data}(x,y)} [\log(1 - D(G(E(x),y)))]\}\} \quad (7)$$

The generator G functions as a decoder for decoding the feature vector $E(x)$ and the condition vector y generated by the convolutional encoder E into a composite image. One of the goals of generator learning is to minimize the probability that the discriminator discriminates as a composite image. According to formula (5), the generator's adversarial loss is expressed as:

$$L_g = E_{x,y \sim p_{data}(x,y)} [\log(1 - D(G(E(x),y)))] \quad (8)$$

In order to produce a clearer and more realistic face image, the objective function of the aging synthesis method in this chapter adds three loss terms in addition to the original objective function of GAN: Feature space loss, pixel space loss, and age loss. Enriching the objective function from multiple angles helps to generate a face-aged image with more realistic and clearer details.

1. Feature space loss

The feature space loss represents the difference between the real face image and the synthesized face image on the network feature map. The network here can be either a trained network, such as AlexNet, or a part of the discriminant network. In order to avoid increasing the network complexity, the output of the middle layer of the network is determined as the feature space of the input image, and the loss of the real image and the composite image in the feature space is calculated. Define feature space loss:

$$L_f = E_{x,y \sim p_{data}(x,y)} [\|D'(G(z,y)) - D'(G(x,y))\|^2] \quad (9)$$

wherein, D' is the middle layer network of the discriminator D , z is the feature vector output by the convolutional encoder, and y is the constraint condition, which represents the face image synthesized by the generator. Combined with formula (3), it can be written as:

$$L_f = E_{x,y \sim p_{data}(x,y)} [\|D'(G(E(x),y)) - D'(G(x,y))\|^2] \quad (10)$$

2. Pixel space loss

The pixel space loss is the difference between the real image and the composite image at the pixel level, and the similarity of the two images at the pixel level is constrained, which is defined as:

$$L_p = E_{x,y \sim p_{data}(x,y)} [\|G(E(x),y) - x\|^2] \quad (11)$$

3. Age loss

The age loss represents the difference in age estimates between the real face image and the synthetic face image, and the introduction of age loss causes the generator to synthesize a reasonable age image. Defined as:

$$L_a = E_{x,y \sim p_{data}(x,y)} [||l' - l''||^2] \quad (12)$$

where l' represents an estimated value of the real image in the trained age estimation model, and l'' represents an estimated value of the composite image.

In combination with the above loss term, for convolutional encoder E and generator G , the objective function becomes:

$$\min_{E,G} (\lambda_g L_g + \lambda_p L_p + \lambda_a L_a + \lambda_f L_f) \quad (13)$$

Among them, λ_g , λ_p , λ_a , and λ_f are weight factors of generator countermeasure loss, pixel space loss, age loss, and feature space loss respectively.

4. Experiment

4.1. Dataset

In order to train an efficient aging synthesis model, one of the key elements is to synthesize visual real and reasonable aging face images and to collect enough age images. The training samples in this section collect data sets with age labels, including Morph-II [17], CACD [18], IMDB-WIKI [19], which are currently available publicly. The Morph-II dataset includes more than 15,000 age-and gender-labeled face images with an average of four images per person, but the average age interval between images is 164 days, and the Morph-II data set does not contain face images of people under 16 years old; the CACD dataset ranges from 16 to 62 years old, including about 16,000 face age images of 2000 persons; the IMDB-WIKI data package has over 520,000 face images with age and gender tags and is a very large publicly available data set. Although the age tags of CACD and IMDB-WIKI are not very accurate, they can still be used in the research of face-aging synthesis. In addition, due to the extreme lack of facial images of people under 15 and over 65 years old in the above datasets, this section also collects a large number of images of infants, children, and the elderly from the Internet, in order to balance the proportion of age and sex in the training dataset. Since the collected images are different (size and format of images, multiple faces in one image, etc.), all images need to be pre-processed. We use the face detection technology, Dlib [20], to crop and align the face images. For crawled images, the age labels are given by an age estimator referring to a variety of age estimation algorithms [21–24]. All images are divided into eight categories, i.e., 0–10, 11–20, 21–30, 31–40, 41–50, 51–60, 61–70, and 71–80. Each category contains about 3500 samples with a similar sex ratio.

Some examples of original images are shown in Figure 6. There are people of different races and colors in the world. At the same time, there are many studies on facial biometric recognition. Some studies have made good progress on age, race, and gender feature recognition [25]. It should be pointed out that, regardless of age estimation or aging synthesis, the subjects in this paper actually include people with different skin colors; however, the target race is Westerners because open source databases such as FG-NET contain mainly many facial images of Western countries.



Figure 6. Cross-age original facial images of three people.

4.2. Network Training

The input of the network is a 128×128 pixel RGB face image. The convolution operation basically uses a 5×5 convolution core with a convolution step of 2. In order to accelerate the convergence speed, the input image, age, and gender conditions are normalized to the range $[-1,1]$. Similarly, because the Tanh activation function is used in the output layer of convolutional encoder and generator, the output is also between $[-1,1]$.

The pixel values of the cropped images and the age label vector, which is an eight-dimensional one-hot vector, are normalized to $[-1,1]$. The encoder E maps the input images to the latent vector z , and then z and y are input to G to synthesize a photo-realistic face image, after which, D discriminates real images from generated images. The optimizer of DCGAN adopts the Adam optimizer ($B1 = 0.5$, $B2 = 0.999$), where the learning rate is 0.0002 and the batch-size is 64. Meanwhile, E_y , E_z , G , and D are updated alternatively.

Our experimental platform is: GPU: Quadro P6000, memory: 24G. TensorFlow, developed and maintained by Google Brain, a team of Google Artificial Intelligence, is a symbolic mathematical system based on data flow programming, which is widely used in the programming of various machine learning algorithms. We use the tensorflow framework for coding.

Every 10 epochs are trained, the model is saved once. On the one hand, it is used to test the training effect. On the other hand, it is used to resume on-site training when the training is interrupted unexpectedly. Probably 6 h are needed for training 60 epochs. The generated images of diverse age groups are illustrated in Figure 7.

Due to the low number of training rounds or the influence of face posture, illumination conditions in the input image, and the limited generalization ability of the network itself, the images generated by the generator may be distorted, resulting in a poor aging synthesis effect.

Figure 8 demonstrates some examples of unsatisfactory synthesis.

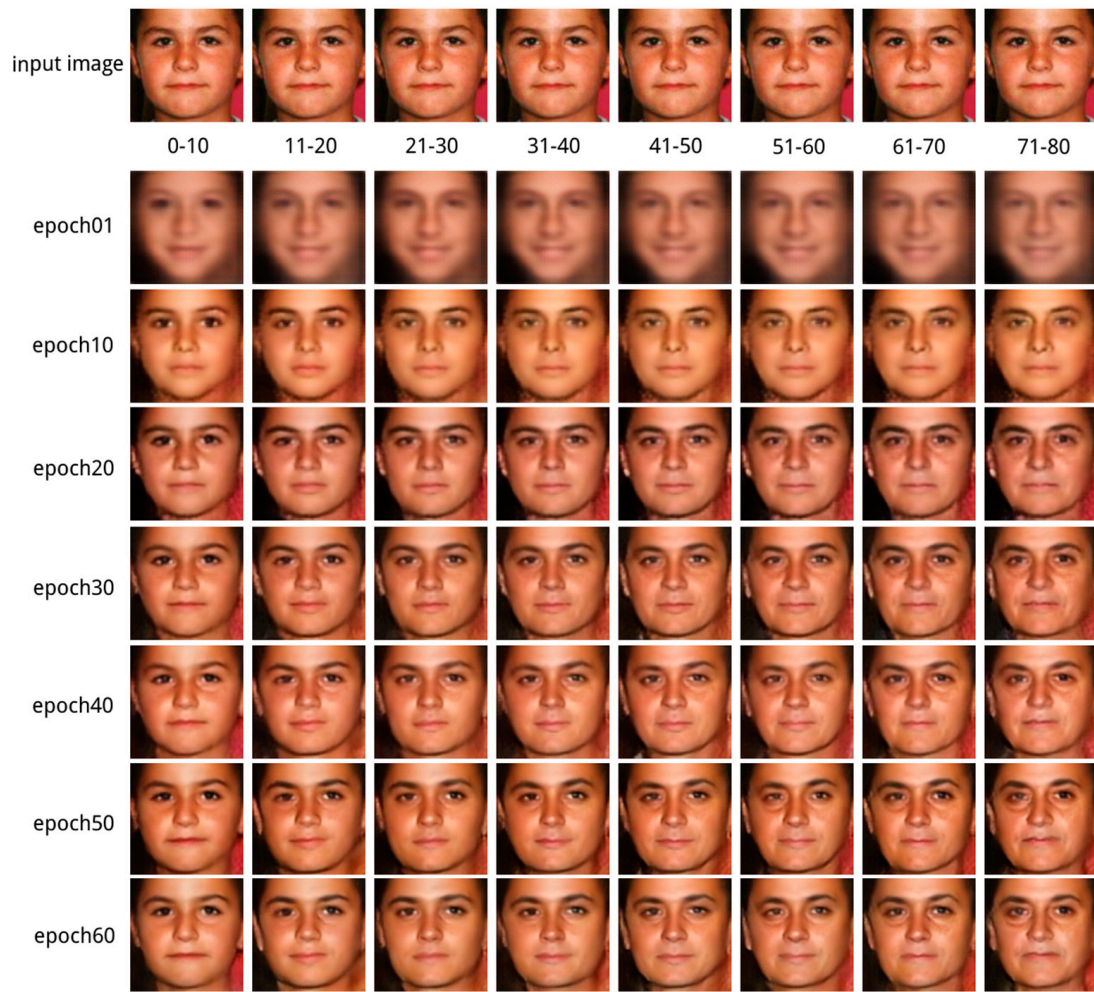


Figure 7. The Intermediate Model Generated by Iterative Training of Different Numbers and the Aging Synthesis Chart Generated by the Model.



Figure 8. Some errors or fails produced by the network.

4.3. Comparison and Evaluation

In fact, it is difficult to judge face-aging methods quantitatively. In this section, we will compare and evaluate our methods from four aspects: robustness, authenticity, comparison with other methods, and age accuracy. We have done a lot of comparative experiments in order to get detailed comparative results in each aspect.

After about 60 epoch training, we can get a better aging synthesis effect. We use the model saved at this time to compare and verify. We use tensorflow to load the trained model and send the input image into the network after pretreatment to get the composite image. Next, we will introduce in detail the comparative experiments of various indicators.

- Robustness

In order to evaluate the robustness of the trained aging synthesis model, we draw some images with graffiti on FG-NET data set and then input them into the aging synthesis model at the same time as the original image. The synthetic images of each age group are obtained respectively, as shown in Figures 7 and 8.

Our input is four pairs of eight face images, each pair of images including an original image and a graffiti picture. By inputting them into our trained generation model, we can get four groups of result diagrams. Each group was composed of original and graffiti aging maps in eight age groups. Although the synthesized face images are not identical to the original synthesized images, they can still obtain better synthesized results.

Eight input drawings which contain four original drawings and four graffiti drawings are shown in Figure 9.

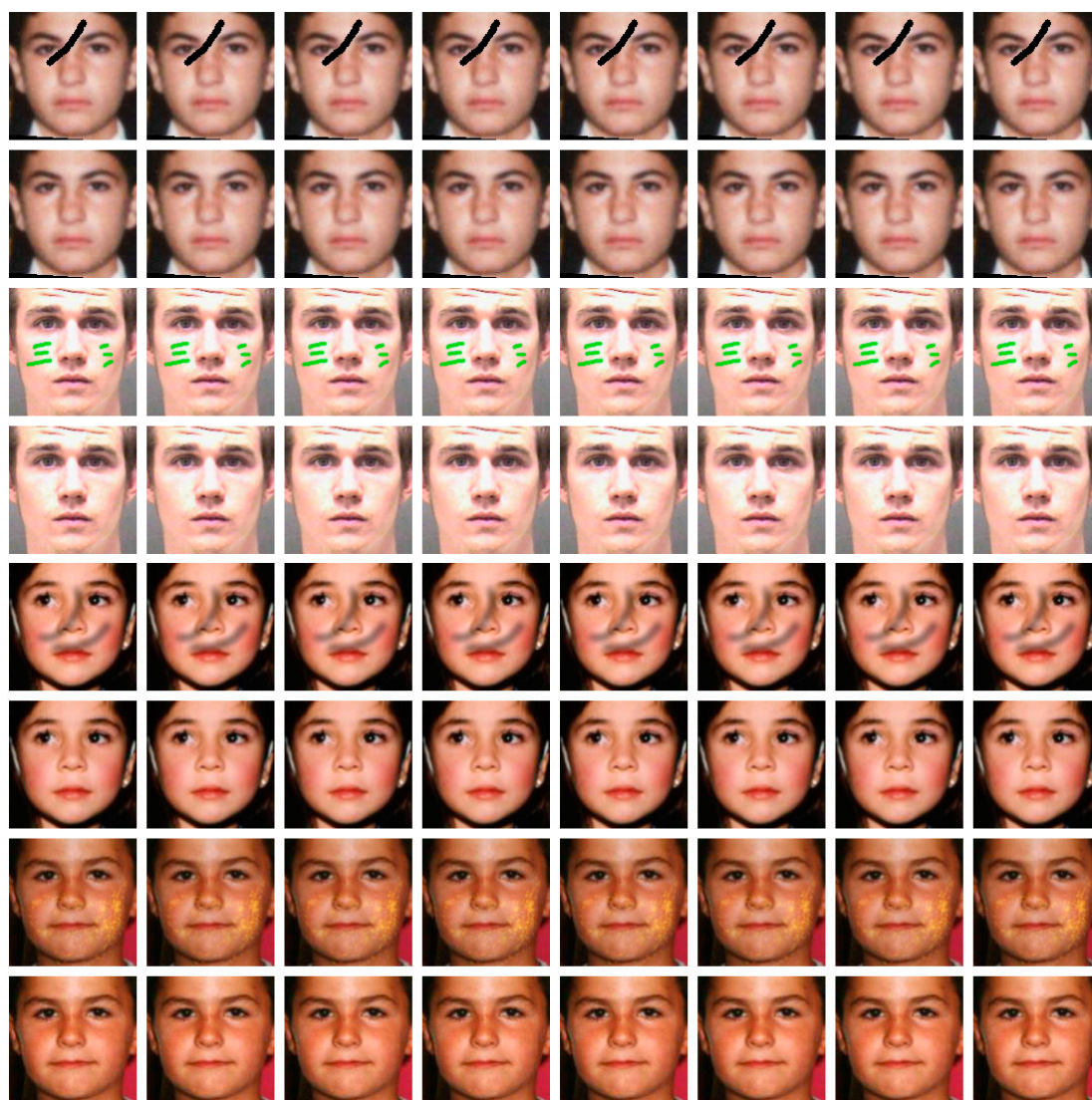


Figure 9. Original graph and pictures with graffiti.

As can be observed, some of the input pictures are covered by black lines, others are covered by markers of other colors. In short, the manner of the graffiti varies. Through the observation of the resulting graph, we can see that the original picture is very similar to the graffiti-generated aging composite map. Generally speaking, the effect is very good, even if some graffiti around the eyes lead to the appearance of glasses in the resulting graph, which is an interesting phenomenon.

The results of the robustness test on FG-NET dataset show that the aging synthesis model constructed by the aging synthesis method in this paper has superior robustness to face occlusion changes.

Eight output drawings which contain four results of original drawings and four results of graffiti drawings are shown in Figure 10.



Figure 10. Synthesis diagrams generated from our model.

- Authenticity

The FG-NET Face Data Set contains 1002 images of 82 people aged 0 to 69. We compared these pictures over the age of 10 years and obtained their age pictures using the synthetic method detailed in this paper. A simple vote is designed to provide participants with three face images, an original A image, a B image generated by our method, and a real image C in the same age group as B. Participants chose one of the three options based on the appearance of A and B: “B is similar to C, B is not similar to C, and is uncertain”. The results of this survey are as follows: 46% of people think that B is similar to C, 19% think that B is not similar to C, 35% feel that it is difficult to judge by the influence of posture and light. Several comparison groups are illustrated in Figure 11, which shows that the method in this paper can better preserve personality and authenticity.

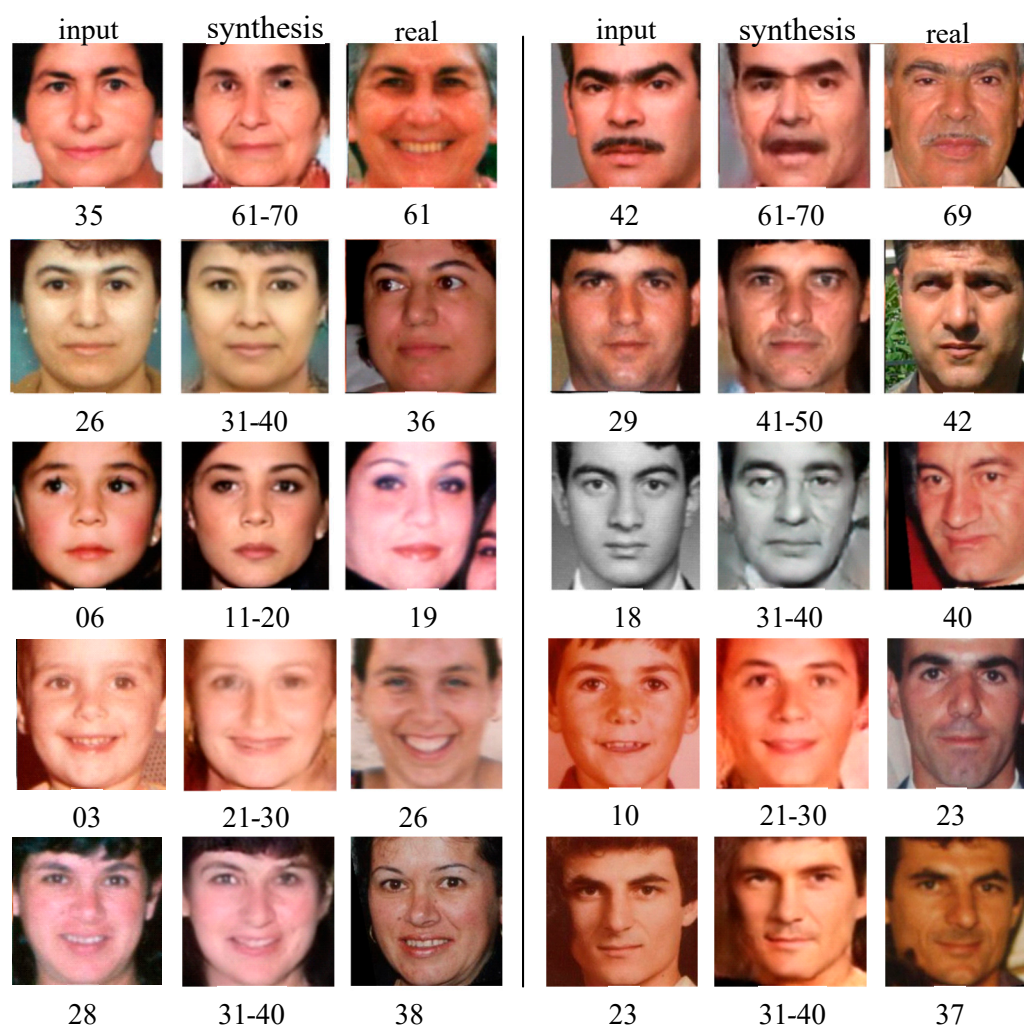


Figure 11. Compared with the real image to judge authenticity.

OpenFace software [26] is an open source face recognition system, we use it to test the input image and synthetic image and judge whether the two faces belong to the same person. There are 2000 pairs of synthetic images and input images totally. We obtain the final result to be 85.25%, which proves the method in this paper can better preserve personality.

- Comparison with prior work

We also use FG-NET face image dataset as the input image and compare it with the Face Transformer Demo (FT Demo) [27]. The comparison results are illustrated in Figure 12, from which we can see that the image generated by FT Demo has obvious ghost defects; especially, the result in infancy has been seriously deformed, where color disorders have occurred. On the contrary, our method can better retain the personality characteristics and generate more realistic and reasonable images.

In addition, we compare our results with several prior face-aging results [7,9,16] and collect some aging synthesis images from published papers. A total of 186 aging images of 56 people were obtained and their corresponding aging faces were generated using the method in this paper. In addition, we conducted the user study that offered four options to the inquirer. It specifically referred to: A is better; B is better; A and B are equal; neither A nor B is good.

These are the statistical results: 39% prefer the method in this paper, 25% consider the prior work to be superior, 19% think they are similar, and 17% think neither is good enough. Several comparison groups are shown in Figure 13. Generally speaking, the method in this paper can generate authentic and reliable aging faces.

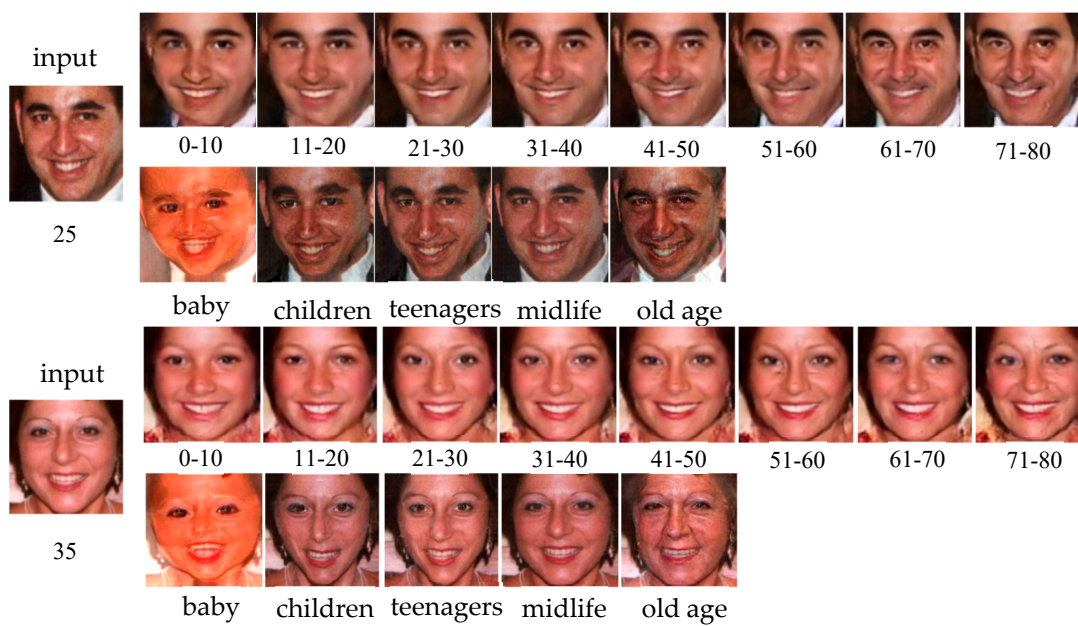


Figure 12. Compared with the Face Transformer (FT) Demos.

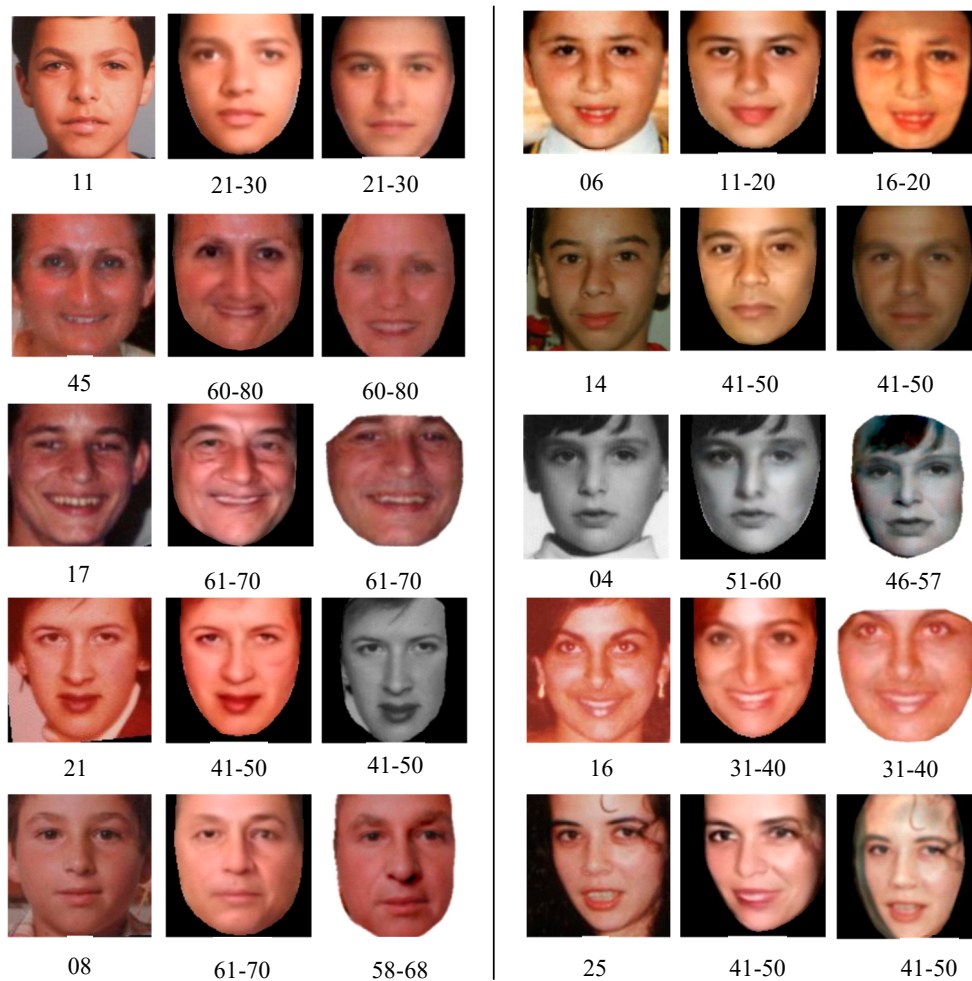


Figure 13. Compared with prior works.

In addition, we also compare with the new results. Shaobo Guan [28] provides a novel method Transparent Latent-Space GAN (TL-GAN) to control the generation process of an unsupervised-trained generative model like GAN (generative adversarial network). Advantages of this method are efficiency and flexibility. The demo can randomly generate a face image, then click on the right button space to select the changed features. There are many optional features, such as hair, eyes, smile, and age, as shown in Figure 14.

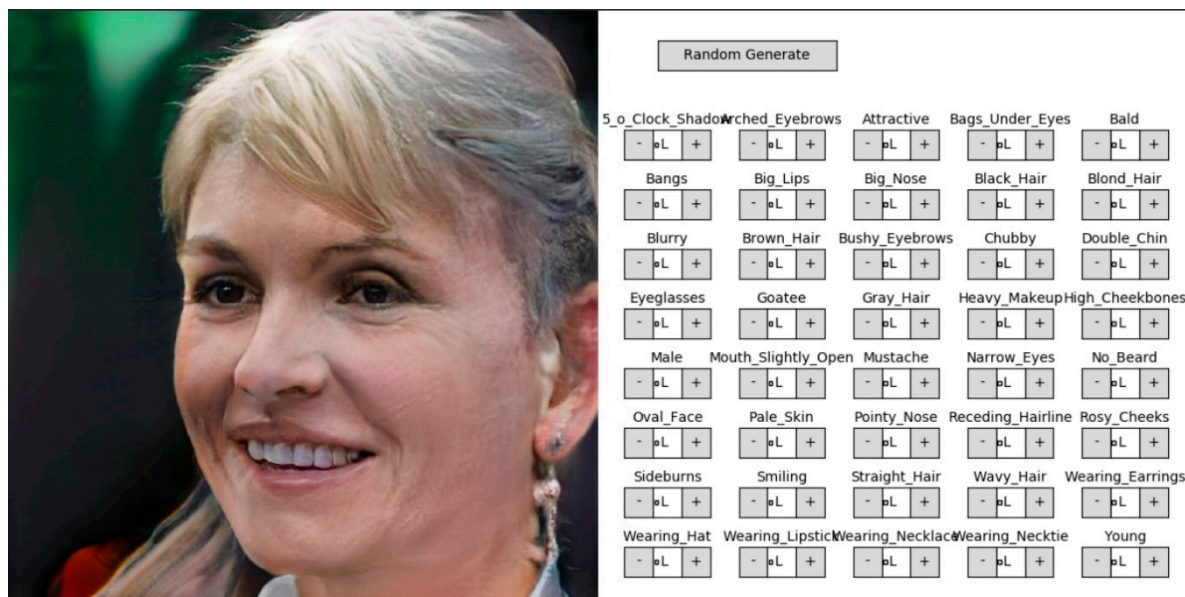


Figure 14. Demo of a Transparent Latent-Space Generated Adversarial Network (TL-GAN).

The experimental results of TL-GAN are shown in Figure 15, from which we can see that we only choose to change the age condition and make a random input image become younger; however, the synthesized image is often not very good. For example, the background is distorted, the hair becomes random, and even the gender may change. This is not an ideal effect. However, the advantages of TL-GAN are obvious, such as its flexibility and interesting function. Indeed, there are too many difficulties and problems that need to be overcome in the field regarding facial-aging synthesis. Many researchers are working hard to find new methods to promote this process. We believe that this will lead to better and more effective aging synthesis methods in the near future.

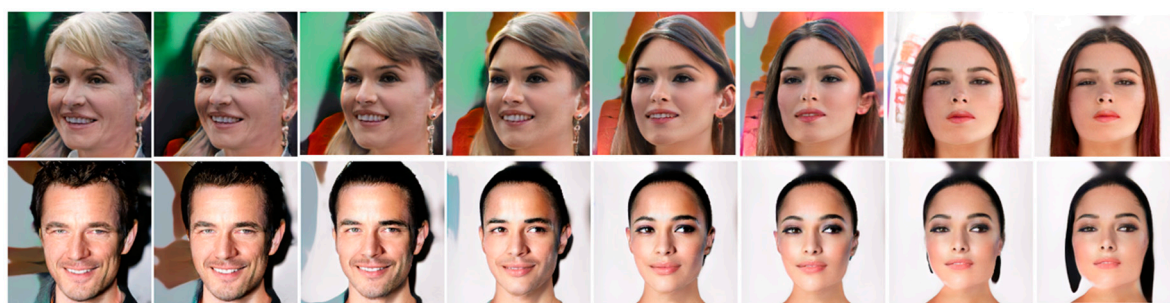


Figure 15. Synthetic effect of TL-GAN.

- Age accuracy

Let us evaluate the reasonableness of age in this section, for example, we generate images of 30–40-years old by setting age conditions, but if we use age estimators to determine their age, the generated images will stand the test. We use the most advanced age estimation CNN to estimate

the age of the generated face. A total of 800 facial pictures were randomly selected from the test data set (100 for each age group), and the gender ratio was adjusted as evenly as possible. When these faces were input into our network, 6400 face pictures were generated. As a contrast, we also selected 6400 real face pictures from the collected data set, with the final test data consisting of 12,800 face images. For each image, the CNN estimates the image based on the age estimates and calculates the final results. The comparison results are illustrated in Table 1. The results show that the accuracy of the generated face is just 8.2% lower than that of the real face, which shows that the method in this paper can synthesize reasonable aging pictures of different age groups.

Table 1. Comparison Result of Age Estimation.

Target	Generated Images	Real Images
Accuracy Ratio	83.6%	91.8%
MAE Error	0.2812	0.1263

5. Conclusions

This paper proposes a simulated face-aging method based on the Deep Convolution Generation Adversarial Network (DCGAN), which preserves the personality characteristics of the face well and has good robustness. Different from the traditional method [29,30], we separate the personality characteristics from the age and gender conditions and map them to low-dimensional vectors through the encoder. At the same time, we apply the perceptual similarity loss to facial-aging simulations to generate photo-realistic aging face images. The rationality and effectiveness of the method are proved by comparing several aspects of the robustness, authenticity, and age accuracy with other methods. The method described in this paper can achieve good age progression and rejuvenation effects. Experiments on several open-source datasets show that the aging synthesis method proposed in this paper can preserve the facial features of the human face well while at the same time generate aging features consistent with the target age range and synthesize visually realistic and reasonable aging images. Most of the aging synthesis methods are based on two-dimensional images; therefore, the next step should be extended to three-dimensional images.

Author Contributions: In this work, X.L. conceived the face aging method based on DCGAN and designed the experiments; Y.Z., C.X. and H.K. performed the experiments and analyzed the data; Y.Z. and C.X. drafted the manuscript; X.L., H.K. and X.M. edited the manuscript.

Funding: The work was supported by the National Natural Science Foundation of China (Nos. 61772088 and 61502361), the Natural Science Foundation of Hubei Province (No. 2014CFB869) and the Fundamental Research Funds for the Central Universities (No. 2014-IV-136).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lanitis, A.; Taylor, C.J.; Cootes, T.F. Toward automatic simulation of aging effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 442–455. [[CrossRef](#)]
2. Fu, Y.; Guo, G.; Huang, T.S. Age synthesis and estimation via faces: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1955–1976. [[PubMed](#)]
3. Dosovitskiy, A.; Brox, T. Generating Images with Perceptual Similarity Metrics Based on Deep Networks. 2016. Available online: <http://papers.nips.cc/paper/6157-generating-images-with-perceptual-similarity-metrics-based-on-deep-networks> (accessed on 17 February 2019).
4. Kang, S.-J. Multi-User Identification-Based Eye-Tracking Algorithm Using Position Estimation. *Sensors* **2017**, *17*, 41. [[CrossRef](#)] [[PubMed](#)]
5. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *arXiv* **2017**, arXiv:1708.02002.

6. Ramanathan, N.; Chellappa, R. Modeling shape and textural variations in aging faces. In Proceedings of the 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, Amsterdam, The Netherlands, 17–19 September 2008; pp. 1–8.
7. Kemelmacher-Shlizerman, I.; Suwajanakorn, S.; Seitz, S.M. Illumination-aware age progression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 3334–3341.
8. Park, U.; Tong, Y.; Jain, A.K. Age-invariant face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 947–954. [[CrossRef](#)]
9. Wang, W.; Cui, Z.; Yan, Y.; Feng, J.; Yan, S.; Shu, X.; Sebe, N. Recurrent face aging. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2378–2386.
10. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
11. Antipov, G.; Baccouche, M.; Dugelay, J.-L. Face aging with conditional generative adversarial networks. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 2089–2093.
12. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
13. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
14. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein gan. *arXiv* **2017**, arXiv:1701.07875.
15. Berthelot, D.; Schumm, T.; Metz, L. BEGAN: Boundary equilibrium generative adversarial networks. *arXiv* **2017**, arXiv:1703.10717.
16. Zhang, Z.; Song, Y.; Qi, H. Age progression/regression by conditional adversarial autoencoder. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4352–4360.
17. Ricanek, K.; Tesafaye, T. Morph: A longitudinal image database of normal adult age-progression. In Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06), Southampton, UK, 10–12 April 2006; pp. 341–345.
18. Chen, B.-C.; Chen, C.-S.; Hsu, W.H. Cross-age reference coding for age-invariant face recognition and retrieval. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 768–783.
19. Rothe, R.; Timofte, R.; Van Gool, L. Dex: Deep expectation of apparent age from a single image. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 7–13 December 2015; pp. 10–15.
20. Dlib C++ Library. Available online: <http://dlib.net/> (accessed on 22 December 2018).
21. Niu, Z.; Zhou, M.; Wang, L.; Gao, X.; Hua, G. Ordinal regression with multiple output CNN for age estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4920–4928.
22. Lu, J.; Liong, V.E.; Zhou, J. Cost-sensitive local binary feature learning for facial age estimation. *IEEE Trans. Image Process.* **2015**, *24*, 5356–5368. [[CrossRef](#)] [[PubMed](#)]
23. Chen, S.; Zhang, C.; Dong, M.J. Deep Age Estimation: From Classification to Ranking. *IEEE Trans. Multimedia* **2018**, *20*, 2209–2222. [[CrossRef](#)]
24. Hu, Z.; Wen, Y.; Wang, J.; Wang, M.; Hong, R.; Yan, S. Facial Age Estimation with Age Difference. *IEEE Trans. Image Process.* **2017**, *26*, 3087–3097. [[CrossRef](#)] [[PubMed](#)]
25. Carcagni, P.; Coco, M.D.; Cazzato, D.; Leo, M.; Distanto, C. A study on different experimental configurations for age, race, and gender estimation problems. *EURASIP J. Image Video Process.* **2015**, *2015*, 37. [[CrossRef](#)]
26. Amos, B.; Ludwiczuk, B.; Satyanarayanan, M.J. Openface: A General-Purpose Face Recognition Library with Mobile Applications. 2016. Available online: https://scholar.google.com.tw/scholar?hl=en&as_sdt=0%2C5&q=Openface%3A+A+general-purpose+face+recognition+library+with+mobile+applications.+&btnG= (accessed on 25 December 2018).
27. Face Transformer (FT) Demo. Available online: <http://cherry.dcs.aber.ac.uk/transformer/> (accessed on 27 December 2018).

28. TL-GAN: Transparent Latent-Space GAN. Available online: https://github.com/SummitKwan/transparent_latent_gan/ (accessed on 17 January 2019).
29. Yang, H.; Huang, D.; Wang, Y.; Wang, H.; Tang, Y. Face aging effect simulation using hidden factor analysis joint sparse representation. *IEEE Trans. Image Process.* **2016**, *25*, 2493–2507. [CrossRef] [PubMed]
30. Shu, X.; Tang, J.; Lai, H.; Liu, L.; Yan, S. Personalized age progression with aging dictionary. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3970–3978.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).