Face Aging Simulation with Deep Convolutional Generative Adversarial Networks

Xinhua Liu[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;
[2]Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Wuhan University of Technology, Ministry of Education, Wuhan 430070, China
Email: xiecj@whut.edu.cn

***Abstract*—Human face aging process simulated by computer has become a hot research issue of the computer vision field. In this paper we propose an improved face aging model based on Deep Convolutional Generative Adversarial Network (DCGAN). In this model, a given face is first mapped to a personal latent vector and age-conditional vector through two sub-encoders. Inputting these two vectors into generator, and then stable and photo-realistic face images are generated by preserving personalized face features and changing age condition. Perceptual similarity loss replace adversarial loss of Generative Adversarial Networks (GANs) as the objective function in this paper. Based on the existing face database, the experiment results demonstrate that face images synthesized by our method enjoys better authenticity and accuracy.**

Keywords- face aging; DCGAN; perceptual similarity loss

I. Introduction

Facial attribute features, especially shape of facial features and facial texture are usually affected by face aging. As the growth of age, one person's facial attributes will change greatly. Therefore, face aging is an inevitable and irreversible change in shape of facial features and facial texture [1]. Technology for simulating face aging process enjoys extensive potential applications including national defense security, information management system, public security, etc. aging simulation technology can enhance the performance of face recognition system, look for missing child and increase the amusement. However, such technology still remains challenging. First, each individual will have unique personality because of complicated facial structure, slow and diversified aging process and various aging reasons. Second, disordered and inconsistent existing face dataset (great difference in facial expressions, postures, occlusion and light conditions). Finally, the existing dataset is hard to meet requirements of various methods. For example, some method requires photos of the same person at different age stages.

To preserve more personality on aging faces and obtain clear ones, we propose an improved age conditional deep convolutional generative adversarial networks framework (Age-DCGAN) in this paper. First, we input the image into the encoder and learn its personality and age features. To learn age feature accurately, we add an age discriminator for age encoder. Then, Age features are input into the generator as the conditions to generate the corresponding aging face. The discriminator imposed on the generator to discriminate input images and generated images. The output of an intermediate layer of the discriminator is selected as image feature that as part of perceptual similarity loss [8].

To summarize, our contributions can be summarized as follows: First, we design two sub-encoders to encode and separate the personality and age features of input facial images, respectively. Second, we apply perceptual similarity metrics to objective function of our model, for acquiring aging faces with clearer outline and details. Finally, our method enjoys great robustness in terms of postures and occlusion.

II. Related Work

A. Face Aging

Traditional face aging simulation methods can be roughly divided into physical model-based methods [2] and prototype-based methods [3]. The physical model-based methods will establish parametric models for facial geometric feature, muscle, wrinkle etc. of human face after fully considering facial features and geometry construction. The prototype-based method divides all the face images into different age groups. It takes the average face of each age group as its prototype, and the difference between prototypes is considered as the aging pattern. However, as aging prototype, the average face will smooth facial texture and cannot capture high frequency details (wrinkles, spots) well. Park, Tong and Jain improved this method and extended it into 3D face aging [4]. 3D face aging simulation method could be obtained by establishing models for shape space and texture space, and certain progress had been made. With extensive application of deep learning in image processing, a simulation method based on recurrent neural networks (RNN) is proposed [5], where RNN is applied for age pattern transition, which can preserve personality better.

Generative Adversarial Network (GAN) has attracted much attention in computer vision field since 2014, especially in image generation, GAN can be trained to generate more realistic faces. Grigory et al. [6] proposed an age conditional generative adversarial network (Age-cGAN) which could preserve recognizable personality on the whole. This is the first time to apply GAN to facial synthesis of given age group. Zhang et al. [7] proposed a conditional adversarial autoencoder (CAAE) that could generate realistic faces by simulating age progression and regression.

B. Generative Adversarial Network

As a generative model, GAN has aroused widespread academic concern since the first proposal by Ian Goodfellow [9]. The original GAN include a generative model which captures the distribution of sample data, and a discriminative model which distinguish the generated samples from the real

2157-1481/18/$31.00 ©2018 IEEE
DOI 10.1109/ICMTMA.2018.00060

ones. And this optimization can be considered as a minimax two-player game as Equation (1), where 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.

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

Fix one model and upgrade parameters of the other model during GAN training, G and D are iteratively trained competing against each other in a minimax game. Then, the distribution of sample data can be assessed by the generative model. GAN is very flexible and generic, and it can integrate different kinds of loss function and avoid the learning mechanism of Markov chain. However there are still some problems, for example, mode collapse may occur during the training process of GAN, and the discriminative model will fail because the generator only generates the same sample points.

During the last two years, various methods have been proposed to improve the original GAN. The Conditional GAN (CGAN) [10] extended the GAN model by adopting additional information to guide the data generated process. DCGAN [11] adopted deconvolutions and convolutions neural networks to enhance its image feature representation capacity. Wasserstein GAN (WGAN) [12] introduced Wasserstein distance, a new way to measure the distance between distributions, which solved the problem of unstable GAN training theoretically. Energy-Based Generative Adversarial Network (BEGAN) [13] combines with WGAN proposed a new equilibrium enforcing method for training auto-encoder based GAN.

III. Proposed Method

A. Model Architecture

The architecture of our face aging model is shown in Figure 1. The framework mainly consists of Encoder (E), generator (G) and discriminator (D). E maps the high-dimensional face image to the latent space and then extracts personality z and age features y' of human faces. The role of E and its detailed structure is given in section B. The Synthetic face $x' = G(z, y')$ will be obtained by inputting z and y' into G. And D will distinguish the generated faces from input faces to generate realistic and reasonable faces.

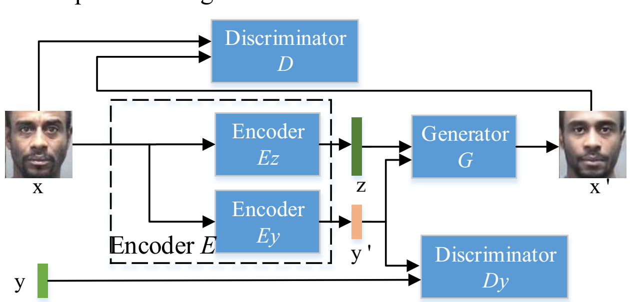

Figure 1. The architecture of our model. A discriminator D_y is added on E_y to discriminate y' and the real age label y.

B. Encoder

Encoder E will be used to map high-dimensional face images into low-dimensional space. Since face aging and personality are diversified, the degree of aging of the same individual at the same age group is also different. To learn age feature accurately, the encoder E proposed in this paper is composed of E_z which maps a face image into $z = E_z(x)$ as personality, and E_y which maps a face image into $y' = E_y(x)$ as age feature. A simple discriminator D_y is imposed on E_y to discriminate y' and real age y, that bring y' approaching y gradually. This is to separate personality from age features, so that we can modify age (while preserving the personality) easily to obtain the reasonable aging faces.

C. Deep Convolutional Generative Adversarial Networks

The part of generative adversarial networks of this paper is based on the implementation of the DCGAN [17], which trains with the Adam optimizer. To generate the face of given age, the age label y' is provided for D as conditional information. In our model, image feature loss L_f and image pixel space loss L_p are added to loss function besides GAN adversarial loss. We do not singly design a CNN network, but adopt the output of an intermediate layer of the discriminator as the image feature. L_f and L_p can be denoted as:

$$L_f = \sum \left\| D'(G(z,y')) - D'(x) \right\|_2^2 , \tag{2}$$

$$L_p = \sum \left\| G(z,y') - x \right\|_2^2 , \tag{3}$$

where D' is an intermediate layer of the discriminator. For GAN adversarial loss, generator are trained by minimizing

$$L_g = -\sum \log(D(G(z,y'))) . \tag{4}$$

Discriminator D are trained by minimizing

$$L_d = -\sum \log(D(x)) + \log(1 - D(G(z,y'))) . \tag{5}$$

The final objective function of network is the weighted sum of these three loss items, D can be trained by Equation (5) and E and G can be trained by

$$Leg = \lambda_f L_f + \lambda_g L_g + \lambda_p L_p . \tag{6}$$

IV. Experiments

A. Data Collection

For the diversity of face data, we collect face images from the public dataset, Morph Aging Dataset [14], CACD dataset [15] and IMDB-Wiki dataset [16]. The Morph dataset which contains more than 13,000 people with 55,000 face images, ranging from 16 to 77 in age. The CACD dataset which includes more than 160,000 face images of 2,000 celebrities. The IMDB-Wiki which contains more than 520,000 face images, is the largest publicly available dataset of face images with gender and age labels. Meanwhile, we crawl a large number of face images from the search engines. Since the collected images are different (size and format of images, multiple faces in one image, etc.), all images need to be pre-progressed. We use the face detection technology, Dilb [17], to crop and align the face images. For crawled images, the age label and gender are given by age estimator [18]. All data 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 has nearly 3,500 samples with similar gender ratio.

B. Implementation

The pixel values of the cropped images and the age label (an eight-dimensional one-hot vector) are normalized to [-1, 1]. The input images through E are mapped to the latent vector z and y' that are input to G to generate a photo-realistic face, and D was used to discriminate real images from generate images. The generated results of different age groups are shown in Figure 2. The implementation of DCGAN, which trains with the Adam optimizer with learning rate is 0.0002 and batch-size is 64. *Ey, Ez, G* and *D* are updated alternatively. About 4 hours are needed for training 60 epochs (GPU: GTX 1070, memory: 8G).

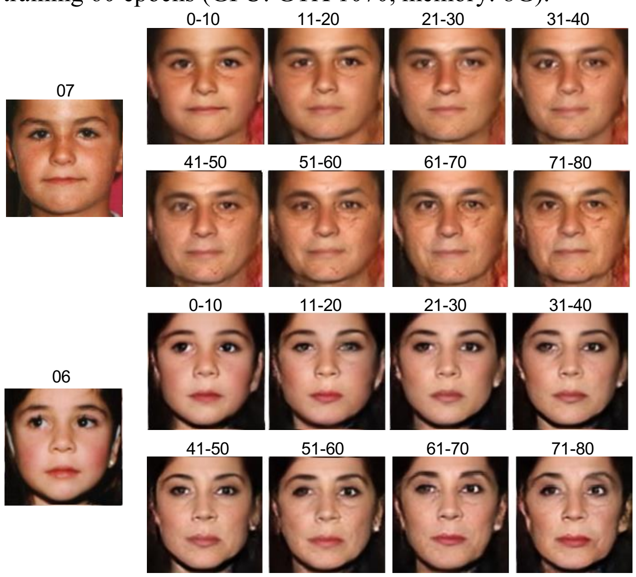

Figure 2. Examples of aging images generated by our model, the first column shows input faces, and the next eight columns are aging faces of eight corresponding age groups.

C. Comparison and Evaluation

Quantitative evaluation: We analyze the performance of our model from two perspectives in this paper. One is the rationality of age. We use the state-of-the-art age estimation CNN to estimate the age of the generated faces. Selecting 400 face images from test data at random, 50 faces (male and female are half each) for each age group. Inputting these faces into our network, 3200 face images are generated. For comparison, we also select 3200 real faces from our collected data, then the final test data consists of 6400 face images. For each image is estimated by the age estimation CNN [18] and the final result will be counted. The comparison result is shown in table 1, which demonstrates that the accuracy on generated faces is only 14.1% lower than on real faces, which indicated that our method can generate reasonable aging images of various age groups.

Table 1 Comparison Result of Age Estimation

	Generated Faces	Real Faces
Accuracy Rate	79.3%	93.4%
MAE Error	0.3114	0.0834

The other is the authenticity of personality. To verify whether the personality has been preserved by input images, we use OpenFace software [19], an open source face recognition system, in our paper to test the input image and generated image and judge whether the two faces are the same person. There are 3000 pairs of generated images and input images in total. We get the final result that is 81.76% which demonstrates our method can preserve personality better.

Comparison with ground truth: the FGNET dataset, which includes 1002 images of 82 people aging from 0 to 69. We select these images with ages range over 10 years for comparison and obtain their aging images by our method. We design a simple vote that provides participants with three faces, an original face *X*, a generated face *Y1* and the ground truth image *Y* in the same age group with *Y1*. Participants select one option from all three options according to whether *Y1* and *Y* look alike. The statistic results of this vote are as follows: 37.62% think the generated face and the ground truth image look alike, 26.92% think they don't look alike, and 35.46% not sure because of the influence of illumination, pose and expression. Some comparison groups are shown in Figure 3 which demonstrates our method can preserve personality and texture better.

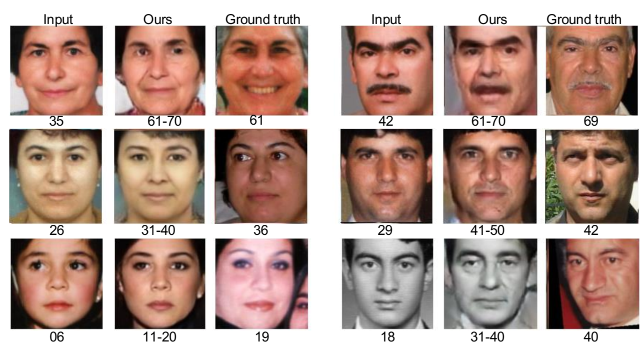

Figure 3. Examples of Comparison with the ground truth.

Comparison with prior work: Face images from FGNET dataset are used as input images to compare with Face Transformer Demo (FT Demo) [20] which is a free online Demo. The comparison result is shown in Figure 4, from which we can see that the images generated by FT Demo present serious ghosting artifacts. Our method can preserve the personality better and the generated images are more photo-realistic.

Furthermore, we compare with some prior face aging results [3, 5, 7, 24, 21, 22] and count some aging synthesis images from their published papers. We obtain 143 aging images of 42 people and then generate their corresponding aging faces by our method. We conduct the user study which offers four options to inquirer (*A* is better; *B* is better; *A* and *B* are equal; neither *A* nor *B* is good). The statistic results are as follows: 32.4% think our method is better, 27.8% think the prior work is better, 32.9% think they are equal, and 6.9% think neither is good. Figure 5 shows some comparison groups. As a whole, our method can generate an authentic and reliable aging face.

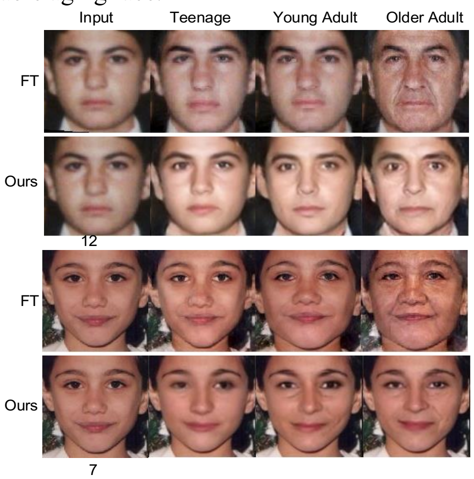

Figure 4. Comparison to FT Demo. The first column is the input faces, and the next columns are aging faces of three age groups: Teenage, Young Adult and Older Adult.

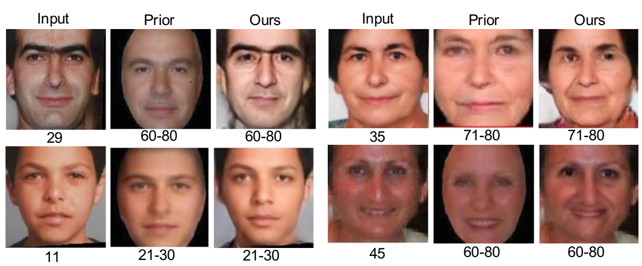

Figure 5. Comparison to prior works. Each row includes an input face and two aging faces of ours and prior works, and the number under the face is age range.

V. Conclusion

An improved age conditional deep convolution generative adversarial networks (Age-DCGAN) is proposed in this paper to simulate the aging process of human face and preserve the personality in a better way. Two sub-encoders are used to separate personality and age features. Meanwhile, we apply perceptual similarity loss to face aging simulation for generating photo-realistic aging faces. Rationality of the method has been proved through analysis and comparison in diverse aspects. Our method can obtain a good effect on age progression, but the effect on age regression is not so good, especially adult male. In age regression work, there are some children and teenagers that are simulated by our method look less realistic because of beard. In the future, an in-depth research on face aging process will be conducted to sharpen facial features and outline, and find more advanced aging synthesis methods.

Acknowledgment

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

References

[1] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," Pattern Analysis & Machine Intelligence IEEE Transactions on, vol. 24, no. 4, pp. 442-455, 2002.

[2] N. Ramanathan and R. Chellappa, "Modeling shape and textural variations in aging faces," in IEEE International Conference on Automatic Face & Gesture Recognition, 2008, pp. 1-8.

[3] I. Kemelmachershlizerman, S. Suwajanakorn, and S. M. Seitz, "Illumination-Aware Age Progression," in IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3334-3341.

[4] U. Park, Y. Tong, and A. K. Jain, Age-Invariant Face Recognition. IEEE Computer Society, 2010, pp. 947-954.

[5] W. Wang et al., "Recurrent Face Aging," in Computer Vision and Pattern Recognition, 2016, pp. 2378-2386.

[6] G. Antipov, M. Baccouche, and J. L. Dugelay, "Face Aging With Conditional Generative Adversarial Networks," in IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[7] Z. Zhang, Y. Song, and H. Qi, "Age Progression/Regression by Conditional Adversarial Autoencoder," in IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4352 – 4360.

[8] A. Dosovitskiy and T. Brox, "Generating Images with Perceptual Similarity Metrics based on Deep Networks," 2016.

[9] I. J. Goodfellow et al., "Generative adversarial nets," in International Conference on Neural Information Processing Systems, 2014, pp. 2672-2680.

[10] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," Computer Science, pp. 2672-2680, 2014.

[11] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," Computer Science, 2015.

[12] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017.

[13] D. Berthelot, T. Schumm, and L. Metz, "BEGAN: Boundary Equilibrium Generative Adversarial Networks," 2017.

[14] K. Ricanek and T. Tesafaye, "MORPH: a longitudinal image database of normal adult age-progression," in International Conference on Automatic Face and Gesture Recognition, 2006, pp. 341-345.

[15] B. C. Chen, C. S. Chen, and W. H. Hsu, "Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval," in European Conference on Computer Vision, 2014, pp. 768-783.

[16] R. Rothe, R. Timofte, and L. V. Gool, "DEX: Deep EXpectation of Apparent Age from a Single Image," in IEEE International Conference on Computer Vision Workshop, 2016, pp. 252-257.

[17] Dlib C++ Library. http://dlib.net/.

[18] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal Regression with Multiple Output CNN for Age Estimation," in Computer Vision and Pattern Recognition, 2016, pp. 4920-4928.

[19] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications." Tech. Rep., CMU-CS-16-118, CMU School of Computer Science, 2016.

[20] Face Transformer (FT) demo. http://cherry.dcs.aber.ac.uk/transformer/.

[21] 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 A Publication of the IEEE Signal Processing Society, vol. 25, no. 6, pp. 2493-2507, 2016.

[22] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan, "Personalized Age Progression with Aging Dictionary," in IEEE International Conference on Computer Vision, 2015, pp. 3970-3978.