Journal of Computers Vol. 34 No. 1, February 2023, pp. 63-73
doi: 10.53106/199115992023023401005

Face Age Feature Analysis Based on Improved Conditional Adversarial Auto-encoder (I-CAAE)

Jia-Li Li[1,2], Xing-Guo Jiang[1,2*], Li He[1,2], De-Cai Li[1,2]

[1] Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Zigong 643000, China
2456761771@qq.com, tonny_jiang@suse.edu.cn
[2] School of Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong 643000, China
1191401486@qq.com, 1340998024@qq.com

Received 7 April 2022; Revised 14 July 2022; Accepted 10 August 2022

Abstract. In recent years, the research of face age features has achieved rapid development driven by deep learning. The faces generated by the Conditional Adversarial Auto-encoder (CAAE) model are not only highly credible, but also closer to the target age. However, there are many problems, such as low resolution of human face image generation and poor local feature retention effect of human face features. To this end, this paper improves on the CAAE network. Firstly, referring to the LSGAN network structure, the 4 convolution layers of the encoder are added to 5 layers and the 4 convolution layers of the generator are added to 7 layers. Secondly, on the basis of the original loss function, the image gradient difference loss function is added to ensure the output face image quality. Meanwhile, the data set were preprocessed for face correction. Finally, this paper performs face similarity analysis on the Eye-key platform and contrasts the generated image quality using structural similarity and peak signal to noise ratio metrics. In addition, the generated results were tested for their robustness. The experimental results show that the average similarity of faces generated by the Improved Conditional Adversarial Auto-encoder (I-CAAE) network was increased by 3.9. And the average peak signal to noise ratio of the generated pictures was reduced by 1.8. Confirming the superiority of the proposed method.

Keywords: age feature, deep learning, conditional adversarial auto-encoder (CAAE), image gradient difference loss function, robustness

1 Introduction

Face image aging research [1], also known as face age feature progression research. It is mainly to combine the geometric features of the face image with the texture features of the age to generate the face image of the specified age. The study of face age feature progression has high application value in the following fields [2]: Firstly, in criminal investigation, it can be used to find missing persons and identify criminals, etc. Secondly, in film and entertainment, it can be used for fancy special effects services and cross-age performance in the film and television industry. Meanwhile, in terms of social life, it can be used for identity document information verification without regular updating of ID document information, saving time and convenience. But, to accurately carry out research on face age, more stringent requirements on the extraction of face age features will be needed. The current research is mainly based on deep learning. The face features are extracted from a large number of data samples for feature representation, and the face change pattern is learned to generate face images of specified age. Among them, the Conditional Adversarial Auto-encoder (CAAE) model [3] is particularly well known. But the research is not yet mature and faced with defects such as poor quality of generated images and poor face personality. For these problems, this paper improves CAAE. Namely, the local feature structure of the output face image is further constrained so that the generated face image has a better local feature structure. Then, the encoder and decoder are further improved to improve the overall effect of face generation for the generated image organ distortion problem.

The contributions are follows as: Firstly, referring the structure of Least Squares Generative Adversarial Networks (LSGAN) [4], this paper adds layers to the generator network and the encoder network to make the generated images keep more face personality features. Secondly, the Image Gradient Difference (IGD) loss function [5] is used to ensure that the generated image contains more high-frequency information, such as the edge contours of the input image, and guarantee the quality of the generated image. At the same time, in order to ex-

* Corresponding Author

tract features better, the experimental data set is pre-processed by face correction, which can improve the accuracy of the model. For narrative purposes, this paper calls this method as Improved Conditional Adversarial Auto-encoder (I-CAAE). Finally, the experimental evaluation metrics were added and compared with CAAE network. The final results show that the I-CAAE network model can generate faces with clearer picture quality and more similar face local feature structure.

2 Related Works

In recent years, the research of face age features has become a hot research topic. The research methods are mainly divided into three different methods [6]: prototype, physical model and deep learning. The prototype-based approach is to synthesize the target face using the prototype face images of different ages and genders [7]. This method lacks face personality expression, and it is easy to appear virtual shadow and artifact phenomenon. The physical-based model simulates the skeletal contour growth and skin texture changes of the face by establishing a parametric model of the face [8]. However, the related parameters of such methods are complex and the calculation process is too relatively complex. With the innovation of deep learning and neural network, the deep learning method applied to the study of face age features has achieved remarkable results.

Dong et al [9] proposed an age-aging model based on a time-depth-constrained Boltzmann machine, which can better capture the nonlinear change process of faces. But the method requires paired training data to ensure that the synthesized face images are of high quality, which has more stringent requirements on the dataset. To solve the problem of difficulty in collecting paired experimental training samples, most of the existing methods are based on Generative Adversarial Nets (GAN) [10] models. The literature [11] combines Auto-encoder (AE) with Conditional Generative Adversarial Nets (CGAN) to implement age-conditional face image domain transformation. Zhang et al [12] proposed a Conditional Adversarial Auto-encoder Network (CAAE), which is one of the best models so far. The generated faces are not only highly credible, but also can be synthesized with a large age span. The model achieves cross-age face synthesis in the overall framework of unpaired input and output images. Compared with the previous one-way face synthesis method, the method can achieve two-way face changes simultaneously. However, this method has defects such as low image resolution and poor local features. Therefore, this paper proposes an improved CAAE network, which can solve the problems mentioned above.

3 Improved Conditional Adversarial Auto-encoder (I-CAAE)

3.1 CAAE Network Model

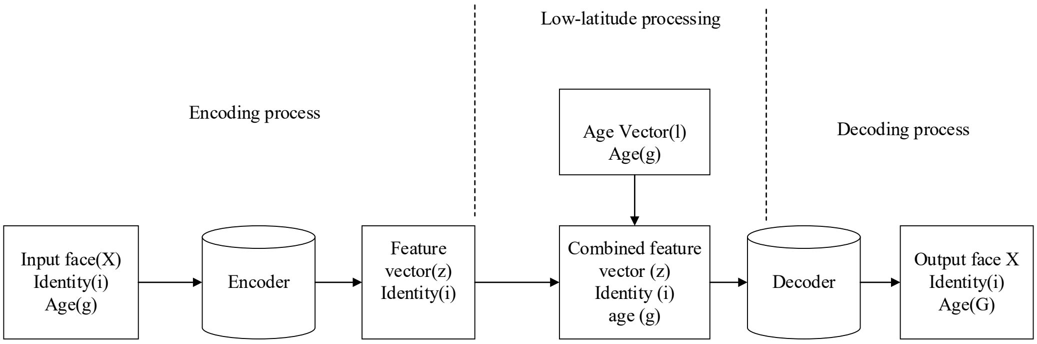


Fig. 1. Principle of image transformation

Conditional Adversarial Auto-encoder (CAAE) model uses the image transformation principle of combining auto-encoder [13] and conditional adversarial network, as shown in Fig. 1. The whole process can be summarized as the image encoding process - low latitude processing - image decoding process.

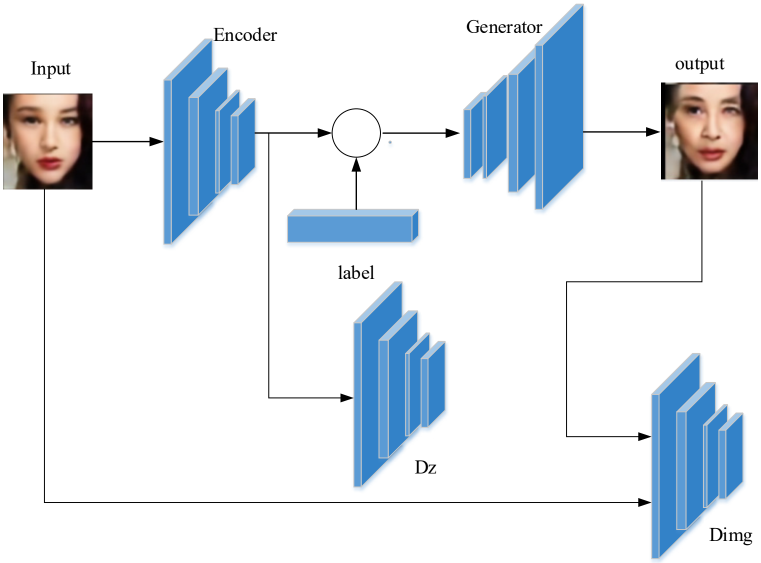


Fig. 2. CAAE network model

The traditional CAAE network model is shown in Fig. 2. As the main structure, the encoder and generator mainly complete the reconstruction of the input face image, and the two discriminators are used to generate more realistic faces. Face images are located in high dimensional space, but the research on face in high dimensional space is relatively complex. So, the low-dimensional features with face identity information need to be obtained, and the low-dimensional features are processed. The operation difficulty is greatly reduced. In the training CAAE network phase, the input face image is mapped to the low-dimensional space by the encoder to obtain vectors with face personalized features. The vectors are assembled with the age label vector, and the assembled vectors with age and face feature information are restored to the high-dimensional space through the generator to obtain the target face image. The discriminant D_z applied on the feature vector produces a more uniform face image by imposing a prior distribution. The discriminator D_{img} makes the generated face image more realistic by minimizing the distance between the input and the output image.

3.2 CAAE Model Improvements

The deeper the network layer, the higher the feature abstraction degree. Therefore, to solve the problem that the local features of face generated by CAAE network model are not well preserved, referring to the LSGAN network structure, the 4 convolution layers of the encoder are added to 5 layers, as shown in Fig. 3. Meanwhile, the 4 convolution layers of the generator are added to 7 layers, as shown in Fig. 4. This paper calls the method as an improved conditional auto-encoder network (I-CAAE).

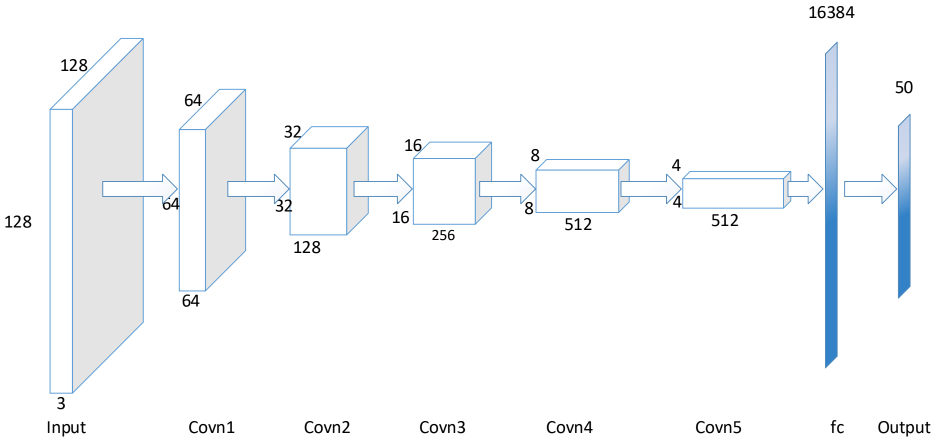


Fig. 3. Encoder E network module

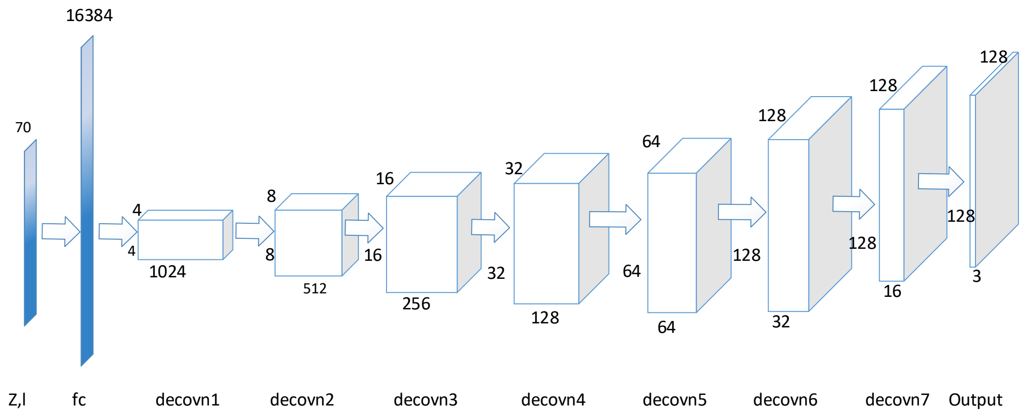


Fig. 4. Generator G network module

In Fig. 3, the function of encoder is to optimize the input picture vector and generate face feature vector after the encoder. The encoder receives the image of 128 × 128 × 3 and passes through five convolutions with core size 5 × 5 andstep length 2, and finally connects to the full connection layer.

In Fig. 4, the generator accepts the vector of the encoder and the age label, and passes through a fully connected layer, and then a deconvolution layers with same size 5 × 5 andstep length 2. Finally the vectors are recovered to the RGB image of 128 × 128 × 3 by using the deconvolution.

3.3 Increase the Loss Function

In order to make the generated images contain more high-frequency information, such as the edge contour of the input image, and ensure the generated image quality. The image gradient difference loss function is added on the basis of the original loss function. The image gradient difference loss function is used to narrow the difference between each pixel of the input and output images, so that the pixel value of each point of the output image is as equal as the input image. The image gradient difference loss function is shown as following

$$Lgdl=(x,\bar{x})=\sum\nolimits_{i,j}\left\|x_{i,j}-x_{i-1,j}\right|-\left|\bar{x}_{i,j}-\bar{x}_{i-1,j}\right|+\left|x_{i,j-1}-x_{i,j}\right|-\left|\bar{x}_{i,j-1}-\bar{x}_{i,j}\right\|. \quad \textbf{(1)}$$

Where, x is the input image, and $\bar{x}$ is the output image, $x_{i,j}$ represents the pixel value of the input image at (i,j) pixel point. The formula takes the pixel as the basic unit, calculates the gradient difference between the input image and the generated image in the direction of row and column, and reduces the difference between them. Discriminator D_z is used to impose constraint on low-dimensional space face features. The loss function of the discriminator D_z is shown as

$$L_{GD_Z}=E_{Z^*\sim P(Z)}[\log D_Z(Z^*)]+E_{x\sim p_{data}(x)}[\log(1-D_Z(E(x)))]. \quad \textbf{(2)}$$

Where, $p_{data}(x)$ represents the distribution of the training data, $p(z)$ is a prior distribution, $z^*\sim p(z)$ shows the random sampling of $p(z)$. The discriminator D_{img} is used to make the synthetic faces more realistic. Its loss function is

$$L_{GD_{img}}=E_{x,l\sim p_{data}(x)}[\log D_{img}(x,l)]+E_{x,l\sim p_{data}(x)}[\log(1-D_{img}(G(E(x,l)))]. \quad \textbf{(3)}$$

Where, $(x, 1)$ shows that face image x age is $l, G(E(x),l)$ represents the face image generated by the generator. $E(x)$ is the mapping vector of the input face through the encoder. The final total loss function is shown as

$$L=\lambda_1 L_{gdl}+L_{GD_Z}+L_{GD_{img}}. \quad \textbf{(4)}$$

Where, L_{gdl} is the image gradient difference loss function, λ_1 is the weight of the image gradient difference loss function, L_{GD_Z} is the loss function of the discriminator D_Z, and $L_{GD_{img}}$ is the loss function of the discriminator D_{img}.

3.4 Adding Evaluation Indicators

To evaluate I-CAAE, face similarity is analyzed through the Eye-key [14] platform. Structural Similarity (SSIM) and Peak Signal to Noise Ratio (PSNR) [15] are used to compare image quality. SSIM measures the similarity of images from three aspects: structure, brightness and contrast. For a given pair of images *x* and *y*, SSIM between the two images is expressed as

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\delta_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\delta_x^2+\delta_y^2+C_2)}. \quad \textbf{(5)}$$

Where, μ_x is the average of *x*, μ_y is the average of *y*, δ_x^2 indicates the variance of *x*, δ_y^2 indicates the variance of *y*, δ_{xy} is the covariance of *x* and *y*. C_1 and C_2 are the constant used to maintain stability. Among them, $C_1=(k_1L)^2, C_2=(k_2L)^2$. *L* is the dynamic range of the pixel value, k_1 and k_2 take the default value, namely k_1 = 0.1, k_2 = 0.03. PSNR is an index to evaluate the quality of images based on the error between the corresponding pixels, the equation is expressed as

$$PSNR=10\log_{10}(\frac{(2^n-1)^2}{MSE}). \quad \textbf{(6)}$$

Where, $MSE=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}(x(i,j)-y(i,j))^2$, denotes the mean square error of the two images. *H* and *W* are respectively the height and width of the images.

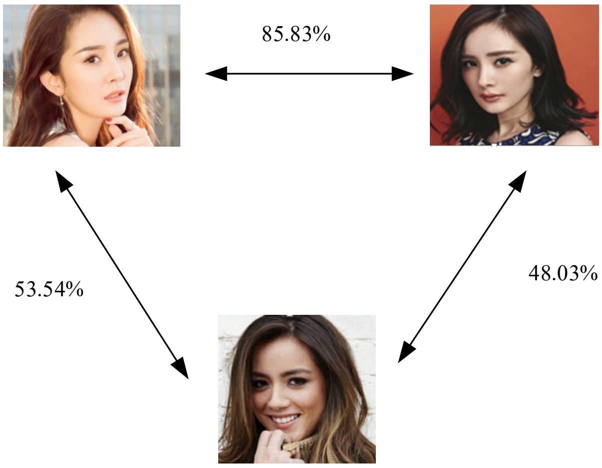
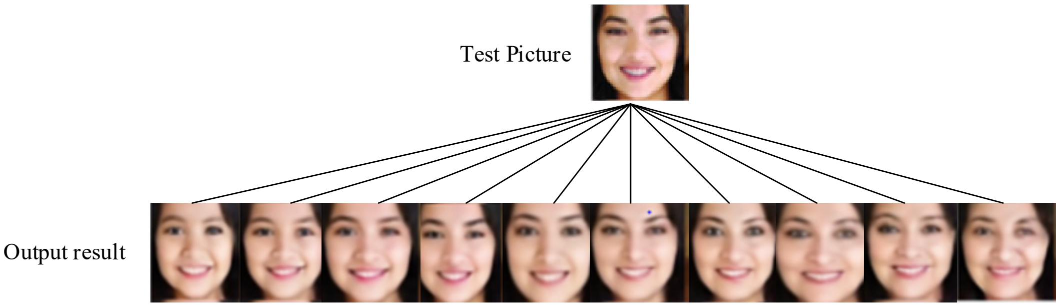


Fig. 5. Face similarity comparison

Fig. 5 shows the relationship of similarity between faces. That is, the faces of the same person have extremely high similarity, while the faces of different people have extremely low similarity.



Similarity contrast	82.68	83.08	85.04	92.12	88.98	88.19	87.41	85.04	85.83	83.46
SSIM	0.43	0.47	0.69	0.68	0.71	0.57	0.68	0.66	0.50	0.34
PSNR	12.17	10.83	15.71	15.11	16.24	12.99	14.90	13.31	13.43	12.42

Fig. 6. Facial structure similarity, SSIM and PSNR comparison

Due to the poor quality of images generated by GAN network, there will be slight blur and shadow phenomenon. So the paper uses SSIM and PSNR indicators to compare the image quality. Larger SSIM values are better, while smaller PSNR values are better.

Fig. 6 shows the comparison of face structure similarity, with SSIM and PSNR between the input and output images. As shown in the Fig. 6, the best structural similarity contrast is 92.15, indicating that the face is most similar to the input face.

4 Experiments and Results

4.1 Experimental Setup

Image Pre-processing. UTK-Face and FG-Net datasets [16] were selected for experiments. The UTK-Face data set contains 23708 face images with age and gender annotation. Download the clipped UTK-Face data set, and select 20000 high resolution faces as the training data set. The FG-Net data set contains 82 photographs of faces in the range of 0-69, as the test data set in this paper.

However, FG-Net data set is not processed by face correction. In order to achieve face alignment, the paper employs the MTCNN algorithm [17] to perform face key point detection and face image orthographic correction. The effect of face correction alignment is shown in Fig. 7.

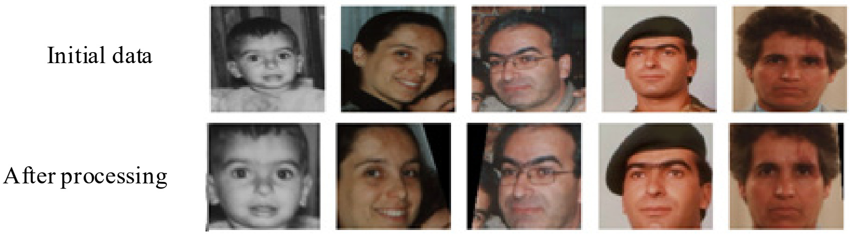


Fig. 7. Face alignment

Experimental Platform and Parameters Selection. The experimental environment is Windows10, 64-bit operating system, Python programming language, TensorFlow deep learning framework, NVIDIA GeForce GTX 1050 Ti GPU and Intel (R)_Core (TM)_ i5-10400F_ CPU.

Through many experiments and comparisons, the parameters of the model are set, as shown in Table 1.

Table 1. Model parameter settings

Parameters	Epoch	Batch	Beta1	Learning rate	Decay rate
Value	25	100	0.5	0.0002	1.0

4.2 Experimental Result

The relationship between the I-CAAE network training loss function and the training times epochs are shown in Fig. 8.

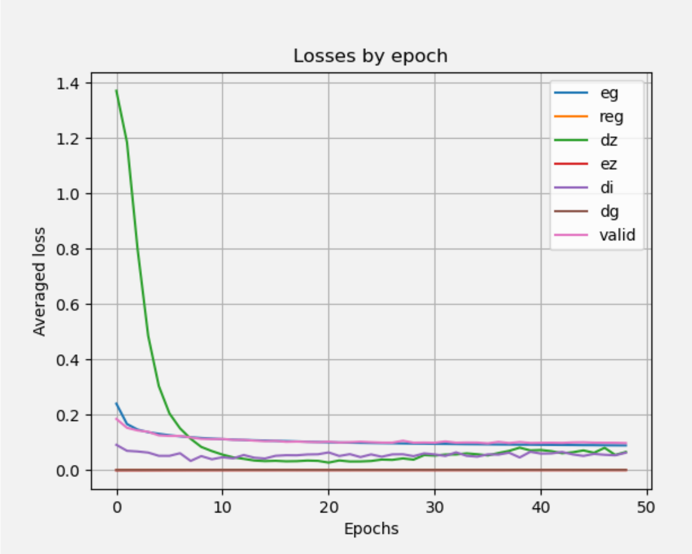


Fig. 8. Loss function

Select faces of different ages for testing. For each face inputted, we obtain the synthetic face images of 0~5 years old, 6~10 years old, 11~15 years old, 16~20 years old, 21~30 years old, 31~40 years old, 41~50 years old, 51~60 years old, 61~70 years old, 70 + years old. Part of the experiment results are shown in Fig. 9.

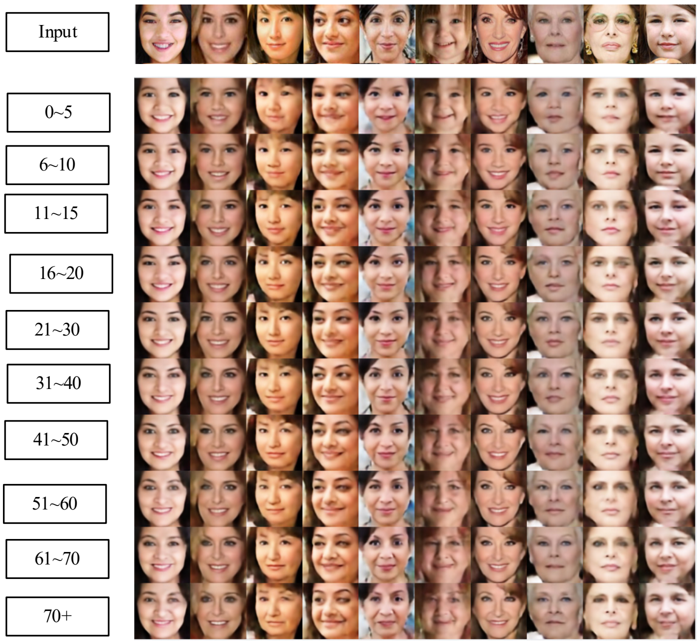


Fig. 9. Generate face result

4.3 Results Analysis

In order to analyze and compare the effect of I-CAAE model with CAAE model, we validate them from qualitative and quantitative perspectives respectively, and add robustness test on the generate result.

Qualitative Analysis. Fig. 10 shows the comparison result of the pictures generated by I-CAAE network and CAAE network. Four groups are shown in Fig. 10. The left side of each group is the face pictures generated by CAAE network, and the right side is the face pictures generated by I-CAAE network. At the same time, for easy contrast, the more different faces are circled with a red box. As can be seen from Fig. 9, the faces generated by CAAE network are no longer obvious with increase of age, and the quality of the generated pictures is fuzzy. While the I-CAAE can well maintain the local features of the face, especially the face wrinkles and changes in the corners of the eyes. Therefore, it can better solve the problems of fuzzy image quality and eyes deformation during synthesis.

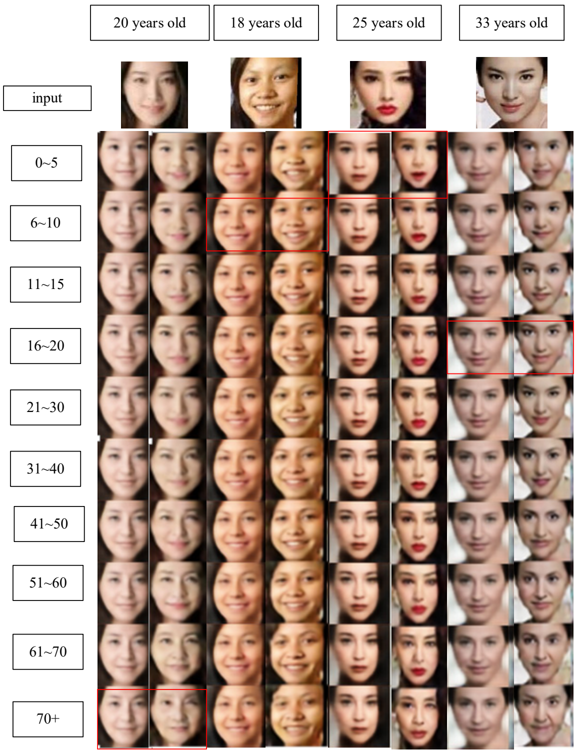


Fig. 10. I-CAAE and CAAE comparison result

Quantitative Analysis. To further verify the effectiveness of the I-CAAE networks, the I-CAAE and the CAAE were then compared from the quantitative indicators, respectively. Two indexes of face similarity and generated photo quality [18] were used for quantitative analysis. First, select 20 faces for testing and the generated faces of each age stage are compared with the test faces. The test results of 20 times are averaged to obtain the histogram shown in Fig. 11. According to Fig. 11, I-CAAE generates higher face similarity in comparison with CAAE, which can indicate that it has better face feature retention effect.

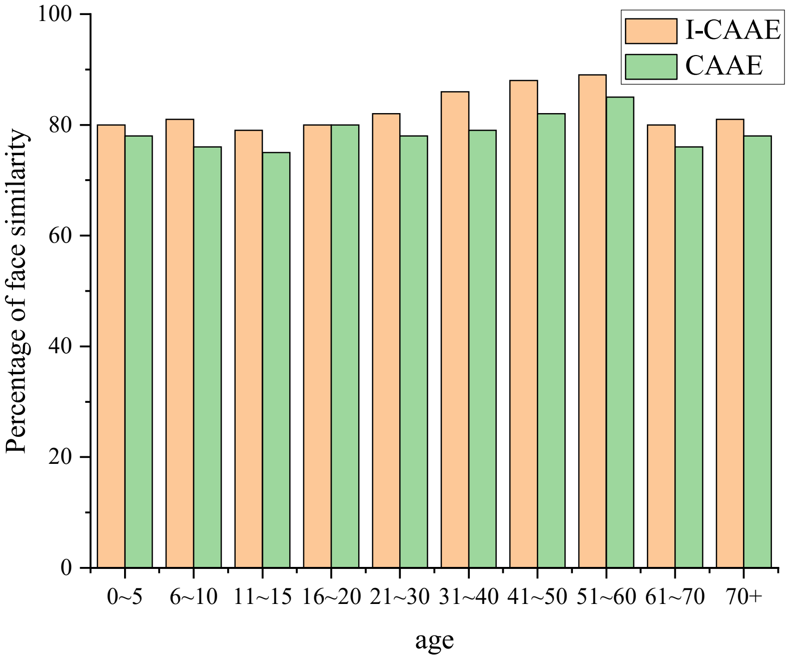
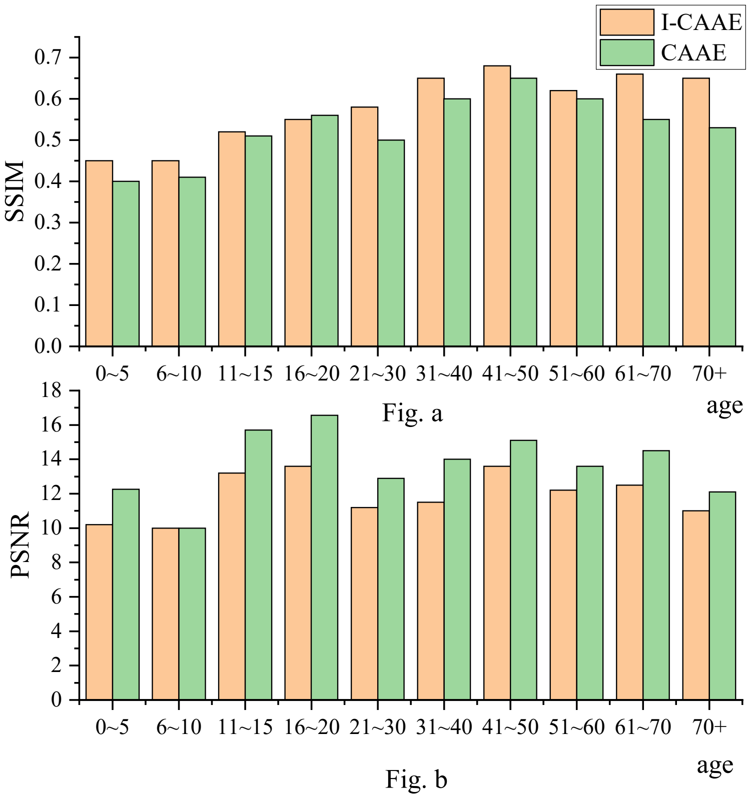


Fig. 11. Face similarity comparison

In addition to the similarity analysis, 20 faces were further selected for SSIM and PSNR testing, and the test results were averaged to obtain the histogram shown in Fig. 12.

Specifically, Fig. 12(a) shows the SSIM comparison result, and Fig. 12(b) shows the PSNR comparison result. As can be seen from Fig. 11, the SSIM of the face generated by I-CAAE is higher, and the PSNR is lower. This further indicates that the quality of face images generated by I-CAAE is better than that of traditional CAAE.



(a) The SSIM comparison result (b) The PSNR comparison result

Fig. 12. Comparison of SSIM, PSNR values

Table 2. Method contrast

	Mirza	Duong	GAN	CAAE	I-CAAE
SSIM	0.78	0.65	0.72	0.53	0.58
PSNR	15.9	15.2	14.8	13.8	11.9

Finally, Table 2 lists the SSIM and the PSNR values for generating pictures of several typical methods. And the results also demonstrate the superiority of the proposed I-CAAE network.

Robustness. To further verify the robustness of the I-CAAE network, the instability of the input face image was simulated by manually adding face mask smearing. Fig. 13 shows the experiment result on face with smearing for CAAE and the I-CAAE, respectively. On the left in Fig. 13, the results are obtained by CAAE, and on the right, the results are obtained by I-CAAE. It can be seen from the results that when the two networks face the same interference occlusion, the faces generated by I-CAAE are more realistic and stronger in face personality, while CAAE network is more serious, and there are face deformations. Therefore, the I-CAAE has stronger robustness.

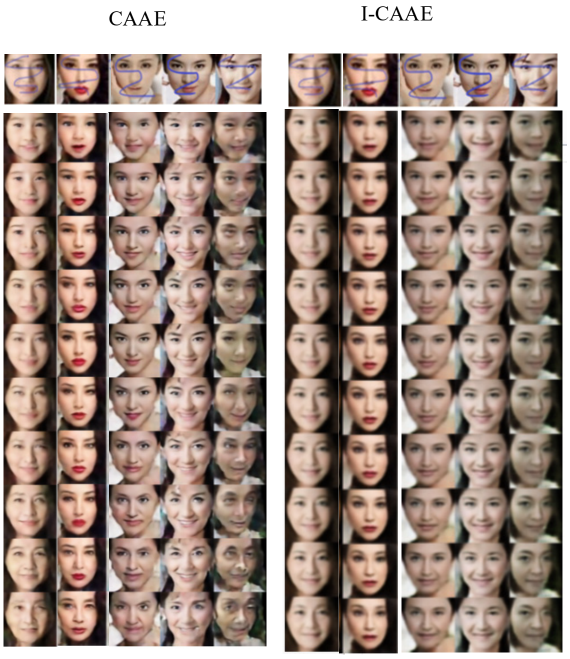


Fig. 13. Robustness testing

5 Conclusions

To address the problems of poor local feature structure retention and low image quality of synthetic images in cross-age face synthesis, this paper proposes an improved conditional adversarial auto-encoder (I-CAAE) network. That is, the encoder and the generator in the CAAE network are processed with additional layers, while the image gradient difference loss function is added. Through comparative analysis on qualitative and quantitative analysis of image quality, we can arrival at a conclusion that the proposed I-CAAE network can improve image quality. Meanwhile, the I-CAAE generates images with richer face feature information and texture information than the CAAE, namely the experiment effect is significantly better than the CAAE. In future study, so that the generated faces have higher image quality. First, we will use more effective loss functions to ensure local face features. And second, find a better generative model to replace the basic Conditions Adversarial Auto-encoder network.

Acknowledgments

This paper was supported by the Open Foundation of Artificial Intelligence Key Laboratory of Sichuan Province under Grant 2020RZJ03, and the Scientific Research Foundation of Sichuan University of Science and Engineering under Grant 2019RC12.

References

[1] T. Wang, M.Y. Zhao, Y.Y. Huang, Overview of Research on Face Age Synthesis Based on Generative Adversarial Networks, Computer Engineering & Software 41(10)(2020) 171-174.

[2] H.Z. Song, X.J. Wu, High-quality image generation model for face aging/processing, Chinese Journal of image and graphics 24(4)(2019) 592-602.

[3] X.H. Wang, H. Lu, X.C. Ma, Generation of Stylized Calligraphic Image Based on Generative Adversarial Network, Packaging Engineering 41(11)(2020) 246-253.

[4] X. Mao, Q. Li, H. Xie, R.Y.K. Lau, Z. Wang, S.P. Smolley, Least squares generative adversarial networks, in: Proc. of the IEEE international conference on computer vision, 2017.

[5] T. Chai, R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)-Arguments against avoiding RMSE in the literature, Geoscientific model development 7(3)(2014) 1247-1250.

[6] H.Z. Song, X.J. Wu, High-quality image generation model for face aging/processing, Chinese Journal of image and graphics 24(4)(2019) 592-602.

[7] A. Lanitis, C.J. Taylor, T.F. Cootes, Toward automatic simulation of aging effects on face images, IEEE Transactions on pattern Analysis and machine Intelligence 24(4)(2002) 442-455.

[8] E. Patterson, A. Sethuram, M. Albert, K. Ricanek, M. King, Aspects of age variation in facial morphology affecting biometrics, in: Proc. 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2007.

[9] C.N. Duong, K. Luu, K.G. Quach, T.D. Bui, Longitudinal face modeling via temporal deep restricted boltzmann machines, in: Proc. of the IEEE conference on computer vision and pattern recognition, 2016.

[10] G. Antipov, M. Baccouche, J.L. Dugelay, Face aging with conditional generative adversarial networks, in: Proc. 2017 IEEE international conference on image processing (ICIP), 2017.

[11] X. Tang, Identity-preserved generative adversarial networks for face aging, Electronic Design Engineering 26(7)(2018) 174-184.

[12] Z. Zhang, Y. Song, H. Qi, Age progression/regression by conditional adversarial autoencoder, in: Proc. of the IEEE conference on computer vision and pattern recognition, 2017.

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

[14] Y.F. Gao, The research of face recognition based on SIFT algorithm under non-ideal condition, [dissertation] Lanzhou: Lanzhou University, 2014.

[15] W. Shi. J. Caballero, F. Huszur, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: Proc. of the IEEE conference on computer vision and pattern recognition, 2016.

[16] X. Bian, J. Li, Conditional adversarial consistent identity autoencoder for cross-age face synthesis, Multimedia Tools and Applications 80(9)(2021) 14231-14253.

[17] J. Xiang, G. Zhu, Joint face detection and facial expression recognition with MTCNN, in: Proc. 2017 4th international conference on information science and control engineering (ICISCE), 2017.

[18] I. Kemelmacher-Shlizerman, S. Suwajanakorn, S.M. Seitz, Illumination-aware age progression, in: Proc. of the IEEE conference on computer vision and pattern recognition, 2014.