IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 June 2023, accepted 10 July 2023, date of publication 12 July 2023, date of current version 9 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3294697

RESEARCH ARTICLE

Facial Expression Transfer Based on Conditional Generative Adversarial Networks

YANG FAN[1], XINGGUO JIANG[1,2], SHUXING LAN[1], AND JIANGHAI LAN[1]
[1]Sichuan University of Science and Engineering, Zigong 643000, China
[2]Artificial Intelligence Key Laboratory of Sichuan Province, Yibin 644005, China

Corresponding author: Xingguo Jiang (tonny_jiang@suse.edu.cn)

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

ABSTRACT With the development of computer vision and image transfer, facial expression transfer has been more and more widespread applications. But there are still some problems, such as lack of realistic expression, poor retention of facial identity features and low synthesis efficiency. In order to solve the problems of facial expression transfer, the paper proposes a facial expression transfer model based on conditional generative adversarial network, which can generate a highly realistic face image with source facial expression and target facial identity features, when gave a source face image and a target face image. The model consists of two parts: the facial feature point fusion module and the expression transfer module. Among them, the facial feature point fusion module uses an auto-encoder to encode the face key feature point image of the source facial expression and the face feature key point image of the target face, so as to transfer the source facial expression information to the corresponding face key feature points of the target image; the expression transfer module uses the facial feature point fusion module to generate the face key feature point image and the target face image, and then generates an image with the source facial expression and the target face identity features through the modified U-net network. The model is finally validated on two publicly available datasets, RaFD and CK+, and the experimental results show that the generated facial expression is more realistic than the pix2pix model, and the model only needs to be trained once to complete the transfer between any facial expression.

INDEX TERMS Face features, facial expression transfer, conditional generative adversarial networks, U-net, face editing.

I. INTRODUCTION

Facial expression is an essential complement to human verbal communication, and rich facial expression can reveal more information, so generating highly realistic facial expression has been a hot research topic in the field of computer vision and image transfer. Among the field, facial expression transfer is one of the important branches, which is to transfer the facial expression of source images to the faces of target images. With the rapid development of intelligent human-computer interaction technology, facial expression transfer is increasingly extensively used in many fields such as data expansion, film and television production and game entertainment [1].

Goodfellow et al. [2] proposed Generative Adversarial Network (GAN) to greatly improve the quality of generated images, but in order to make the generated images have certain desired properties, the literature [3] proposed Conditional Generative Adversarial Network (CGAN), i.e., adding label information to the GAN to constrain the generated images. The literature [4] enabled a single generator to transfer between seven facial expression by specifying the expression domain conditions, and the facial expression generated are more natural and realistic than those generated by the StarGAN model. The paper is precisely based on the conditional generative adversarial network to transfer the source facial expression to the target face using the target face feature

The associate editor coordinating the review of this manuscript and approving it for publication was Michele Nappi.

point images with the source expression as the conditional information and the arbitrary target face images as the input. The main contributions are summarized as follows:

(1) Propose a novel based on conditional generative adversarial network model, including two modules: facial feature point fusion module and expression transfer module. The model can effectively accomplish facial expression transfer.
(2) The autoencoder [5] is used as the facial feature point fusion module to generate the key points of facial features used to guide facial expression transfer.
(3) Modify the structure of the U-net [6] network for the features related to facial expression transfer.

II. RELATED WORK

The current facial expression transfer methods can be divided into two main categories [7]: one is to use the traditional 3D modeling method to build a 3D face parameter model and make the target face match the source facial expression; the other is to use deep learning to generate directly the target face image with source facial expression by generating adversarial network models. Most of the traditional facial expression transfer methods are based on the construction of face models, by mesh fitting and 3-dimensional morphable model (3DMM) to generate new facial expression. Thies et al. [8] proposed Face2Face model, which uses the identity and expression information of each frame to construct a parametric model, and uses the expression parameters of the source face model to replace the corresponding expression parameters in the target face model. Tran and Liu [9] proposed nonlinear 3DMM reconstruction, where the encoder can complete the modeling from 2D to 3D by estimating and double decoding the texture and shape, etc. The literature [10] used an improved parametric-unsupervised regression to transfer the basic expression from the source model to the target model. However, these methods mentioned above need to design different facial model parameters for different characters, thus the generalization is low and the control of facial details is insufficient.

Driven by the proposal and development of deep learning, generative adversarial network (GAN) have shown extremely high fitting ability in image generation and greatly improved the quality of generated images. Pix2Pix [11] changed the input of the original GAN from random noise to specified expression information to complete the transformation between paired images. In facial expression transfer research, researchers have proposed many variants of GANs. G2-GAN [12] used facial geometric information as a conditional vector to guide expression synthesis. However, the model requires the neutral expression image of the person as a medium to achieve expression transfer, which limits the effectiveness of the network to a certain extent. Expr-GAN [13] introduced an expression control module that transforms one-hot vectors with specified expression labels into expression codes to control the generation of face images with different expression intensities. The literature [7], on the other hand, proposed a face synthesis model driven by facial geometric features and attribute labels, and the introduction of a novel soft margin triplet perception loss function allows the model to converge faster and better maintain the identity features of the target face. The literature [14] obtained the respective potential vectors by encoder mapping the face pictures and the key points of face features with facial expression information into the potential space, and the obtained cascaded vectors are then decoded to generate face pictures with the specified expression. The literature [15] removed the batch normalization layer of the encoder and added the expression label feature map to the decoder to make changes to the expression, thus reducing the model parameters to some extent. These aforementioned methods only focus on the transfer of expression to face, while ignoring the identity features of the target faces.

Therefore, to address the above problems, the paper proposes a novel facial expression transfer model based on conditional generative adversarial networks. The model mainly consists of two modules: facial feature point fusion module and expression transfer module. The model can generate a new highly realistic face image with the source facial expression and the target face identity features based on the input source face image and the target face image.

III. FACIAL EXPRESSION TRANSFER BASED ON CGAN

The structure of the proposed model is shown in Fig. 1, which mainly includes the facial feature point fusion module and the expression transfer module. The facial feature point fusion module is denoted as Φ, the expression transfer module is denoted as Ω. The facial feature point fusion module is the part in the dashed box in Fig. 1. Firstly, the feature extractor provided by DLIB is used to pre-train the model for extracting the face feature key points L_s of the source image and the facial feature key points L_t of the target image, and then the features are extracted by encoder G_{Enc1} and G_{Enc2} respectively, and then the two are cascaded and decoded by decoder G_{Dec} to obtain a deviation vector; finally, this deviation vector is fused and summed with the target facial feature point image J_{gen}. The fusion process can be expressed as:

$$\begin{aligned} J_{\text{gen}} &= L_{\text{t}} + \Phi\left(L_s, L_t\right) \\ &= L_{\text{t}} + G_{\text{Dec}}\left(G_{Enc1}\left(L_s\right), G_{Enc2}\left(L_t\right)\right) \end{aligned} \tag{1}$$

The expression transfer module is the modified U-net network in Figure 1. The target image P_t and the generated face feature key point image J_{gen} are fed into the modified U-net network Ω to generate the face image I_{gen} with the source expression and the target face features. The transfer process can be expressed as:

$$I_{gen} = \Omega\left(J_{gen}, P_t\right) \tag{2}$$

A. FACIAL FEATURE POINT FUSION MODULE

The facial feature point fusion module is an auto-encoder network, as shown in Figure 2. The facial feature point

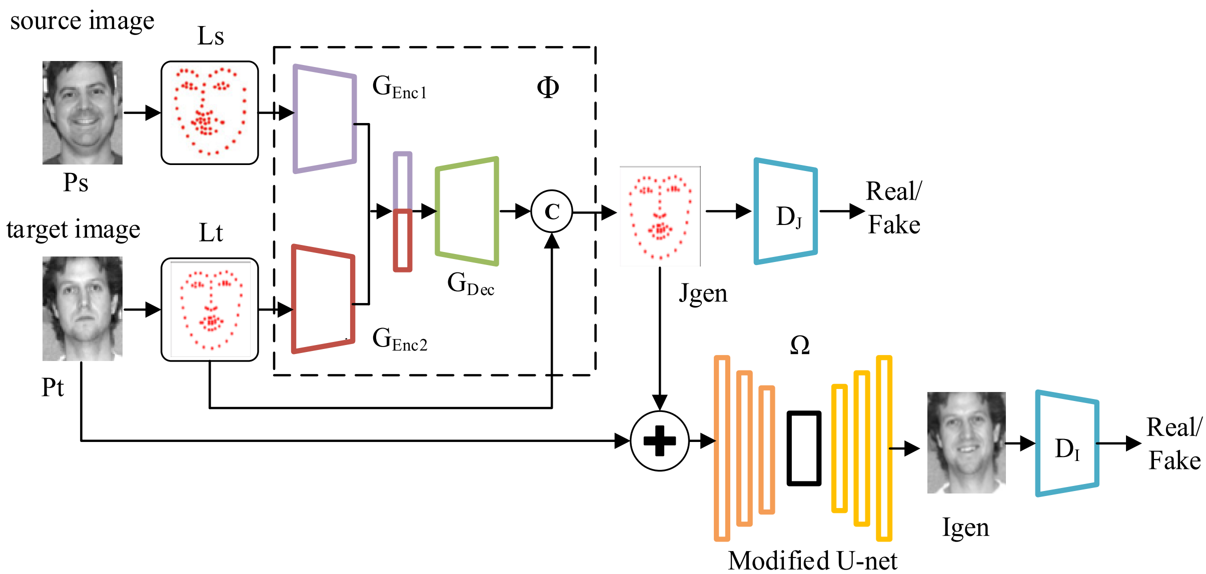

FIGURE 1. Structure of facial expression transfer model.

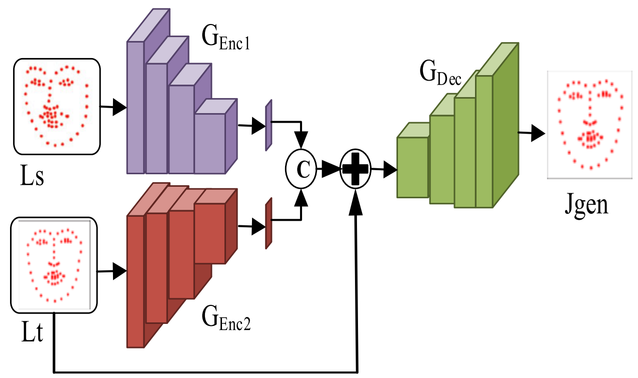

FIGURE 2. Structure of facial feature point fusion module.

fusion module mainly encodes the source face feature key point image L_s and the target face feature key point image L_t. Since the face identity features are different, the same expression is different on different faces, such as the deformation of the senses and the movement of the facial muscles [7].

The facial feature key point information is mapped into a latent space by an encoder, and then the facial feature point fusion module transfers the expression information of any source face to the target face in the latent space (i.e., the position change of the facial feature points).

B. EXPRESSION TRANSFER MODULE

Facial expression transfer is the transfer of only facial expression while retaining most information such as face features and background. Due to the fact that the input source expression image and the target face image are aligned in spatial location, and the encoding-decoding structure and Skip Connection of the U-net network can perform the image transfer well, therefore, the U-net network is modified as the generator of the expression transfer module.

According to DCGAN [16], the convolutional kernel size, activation function and network normalization method of the U-net network are modified as follows:

(1) To make the network itself to learn spatial up-sampling and down-sampling, the network eliminates the maximum pooling layer and uses transposed convolution for the up-sampling process.
(2) To restrict the pixel values of the images to the range $(-1, 1)$, the output layer uses Tanh as the activation function and does not perform batch normalization, and the rest of the network layers use the ReLU activation function.
(3) To avoid the checkerboard artifacts, the modified U-net network uses a difference up sampling and down sampling method, while in order to reduce the model complexity, the convolutional kernel size is setup as 4×4.

The modified U-net network is a fully convolutional network, which consists of three main basic structures: the up-sampling layer consisting of the activation function ReLU, Transposed Convolution with a Stride of 2, and Batch Normalization; the down-sampling layer consisting of the activation function LeakyReLU, Convolution with a Stride of 2, and Batch Normalization; and the bottleneck layer consisting of the activation function LeakyReLU, ReLU, Convolution, and Batch Normalization. The bottleneck layer consists of the activation function LeakyReLU, ReLU, convolution and batch normalization. the input layer is the key point image of the face generated by the target face image and the facial feature point fusion module, and the output layer is composed of ReLU, transposed convolution and Tanh, and the size of both the input and output layers is 256 × 256. The specific network parameters are shown in Table 1.

TABLE 1. Modified U-net network parameters.

Number of layers	Type of layer	Output Dimension
0	Input layer	(256,256,4)
1	Down-sampling	(128,128,64)
2		(64,64,128)
3		(32,32,256)
4		(16,16,512)
5		(8,8,512)
6		(4,4,512)
7		(2,2,512)
8	bottleneck layer	(2,2,512)
9	Up-sampling	(4,4,512)
10		(8,8,512)
11		(16,16,512)
12		(32,32,256)
13		(64,64,128)
14		(128,128,64)
15	Output layer	(256,256,3)

The expression transfer module uses the Markov discriminator, which is a fully convolutional network that can accept inputs of arbitrary size compared to the discriminator of the original generative adversarial network. To improve the stability during training, the batch normalization of the Markov discriminator is changed to Spectral Normalization (SN) to satisfy the Lipschitz continuity condition according to WGAN [17]. The Markov discriminator discriminates true from false for each grid point in the feature map after multi-level down-sampling feature extraction, and the mean value of the feature map of size 30 × 30 is used for the loss calculation.

The Markov discriminator input layer consists of convolution with activation function LeakyReLU (k = 0.2), the down-sampling layer consists of convolution, spectral normalization with ReLU, and the output layer consists of convolution with activation function Sigmoid. The specific network parameters are shown in Table 2.

IV. LOSS FUNCTION

A. LOSS FUNCTION OF THE FACIAL FEATURE POINT FUSION MODUL

Adversarial loss function. It is used to judge whether the generated facial features key points are realistic and reduce the gap between the generated image and the real image, and is defined as:

$$L_{D_J} = E_{x\sim P_{date}(x)}\left[\log D_J(x)\right] + E_{z\sim P_{date}(z)}\left[\log\left(1 - D_J(\Phi(z))\right)\right] \quad (3)$$

TABLE 2. Discriminator network parameters.

Number of layers	Type of layer	Output Dimension
1	Input layer	(128,128,64)
2	Down-sampling	(64,64,128)
3	Down-sampling	(32,32,256)
4	Conv, SN, LeakyReLU	(31,31,512)
5	Output layer	(30,30,1)

where, $P_{data}(x)$ denotes the data distribution of real facial feature points, z denotes a set of inputs to Φ, and D_J denotes the discriminator of the facial feature point fusion module.

Cyclic consistency loss function. It's used to calculate the error between the source facial key points generated by J_{gen} inverse and the real source facial key points, and also allows the network to better preserve the target facial identity features while changing facial expression, defined as:

$$L_{\text{cycle}} = \| \Phi(J_s, \Phi(L_t, L_s)) - L_s \|_1 \quad (4)$$

where, J_s indicates the source facial feature points with reference expression.

L_1 loss function. It is used to calculate the error between the generated target facial key points with source facial expression and the real target facial key points, defined as:

$$L_1 = \| J_{gen} - J_o \|_1 \quad (5)$$

where, J_o indicates the real target face image with the source expression.

Combining Eqs. (3) - (5), the total loss function L_Φ of the facial feature point fusion module is defined as:

$$L_G = \lambda_1 L_{D_J} + \lambda_2 L_{cyc} + \lambda_3 L_{L1} \quad (6)$$

where λ_1, λ_2 and λ_3 indicate the weights of each loss function in the facial feature point fusion module.

B. LOSS FUNCTION OF THE EXPRESSION TRANSFER MODUL

The most significant feature of generative adversarial network is to train the network model through adversarial learning between the generator and discriminator. In order to reduce the gap between generated and real images, and maximize the generation of real and natural images, the expression transfer module uses the adversarial loss function of cGAN, defined as:

$$L_{cGAN} = E_{P_t, I_o}\left[\log D_I(P_t, I_o)\right] + E_{P_t, J_{gen}}\left[\log\left(1 - D_I\left(P_t, \Omega\left(P_t, J_{gen}\right)\right)\right)\right] \quad (7)$$

where, I_O indicates the real target face image with source expression and D_I is the Markov discriminator.

To calculate the pixel-level error between the generated image and the real target image, the expression transfer module also introduces a pixel-level loss function. Isola et al. [11]

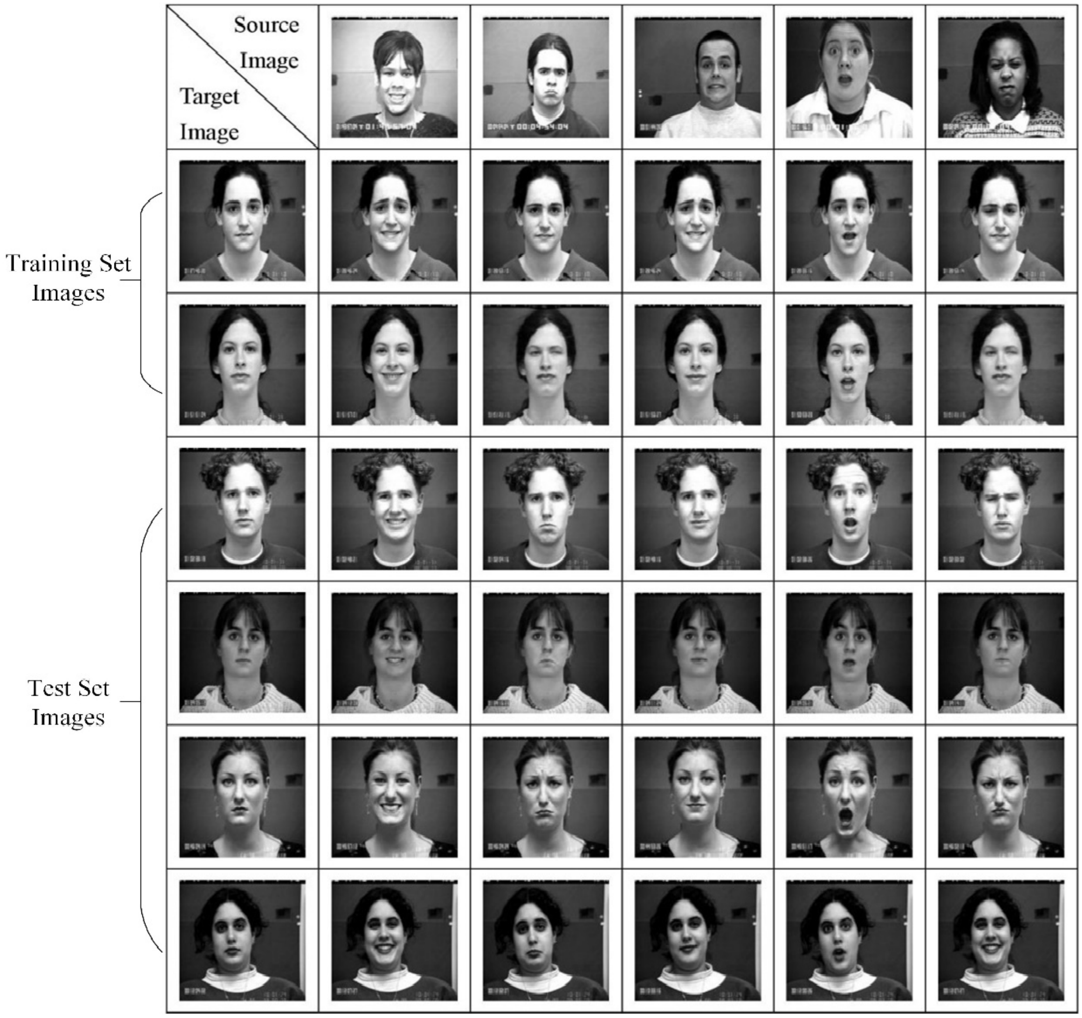

FIGURE 3. Facial expression transfer effect on CK+ dataset.

proposed that optimizing the L_1 parametrization of the error between the generated image and the real image can effectively mitigate the image blur caused by the L_2 parametrization, therefore, the L_1 parametrization is used to define the pixel-level loss function.

$$L_{pix} = \| I_{gen} - I_o \|_1 \tag{8}$$

Combining Eqs. (7) and (8), the total loss function L_Ω of the expression transfer module is defined as:

$$L_\Omega = \mu_1 L_{cGAN} + \mu_2 L_{pix} \tag{9}$$

where, μ_1 and μ_2 indicate the weights of each loss function in the expression transfer module, respectively.

V. EXPERIMENTS AND RESULTS

A. DATASETS

The experiments are conducted using two different types of datasets: the CK+ (The Extended Cohn-Kanade Dataset) [18] and the RaFD (The Radboud Faces Database) [19].

The CK+ dataset contains 123 participants with a total of 593 facial expression sequences. Each sequence started with a neutral expression and ended with a peak expression, and all of them were tagged with the FACS (facial action coding system) coded AU. Most of the images in the CK+ dataset are grayscale, but there are a few color expression sequences, which need to be converted to grayscale before the experiment.

The RaFD dataset contains 67 participants with a total of 8040 images. It contains 8 facial expression, i.e., happy, sad, surprised, angry, fearful, disgusted, contempt and neutral facial expression. Each facial expression contained 3 different gaze directions and was captured simultaneously from different angles using 5 cameras.

Before the experiments started, the two datasets were divided into training sets and test sets in the ratio of 8:2, and the training sets was enhanced with data using a left-right flipping strategy. In addition, the DLIB model was used to extract 68 key points of facial features for each image under the above two datasets.

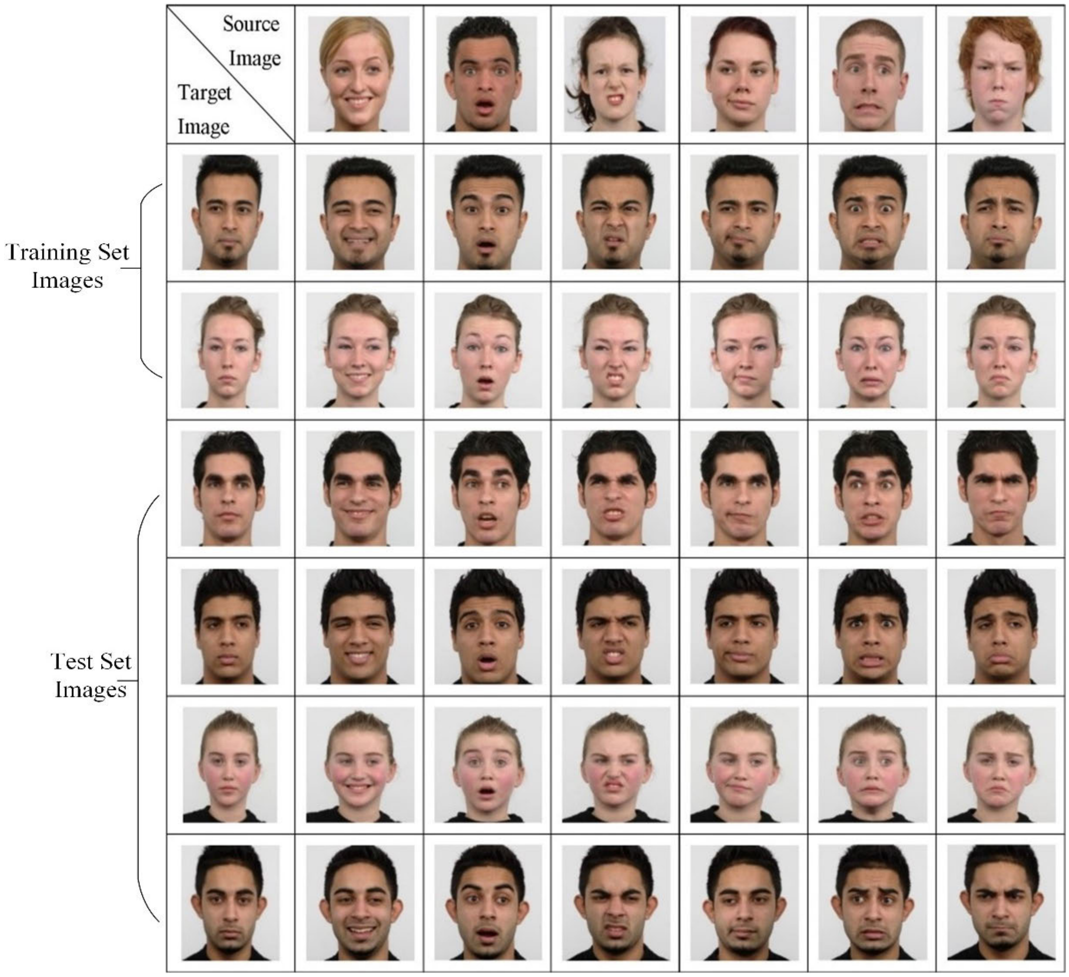

FIGURE 4. Facial expression transfer effect on RaFD dataset.

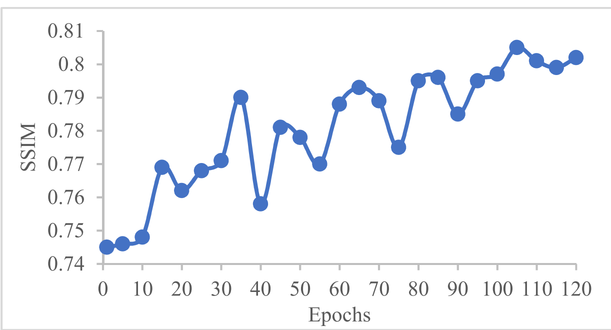

FIGURE 5. SSIM index value change curve with training process.

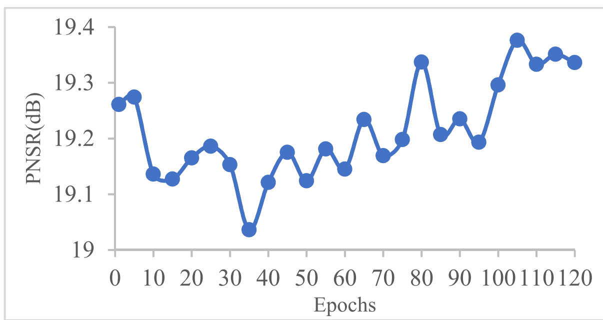

FIGURE 6. PNSR index value change curve with training process.

B. TRAINING METHODS

In the model training, the loss function defined in equation (6) is used to train the facial feature point fusion module first; then the parameters of the completed face feature point fusion module are fixed; and finally, the expression transfer module is trained using the loss function defined in equation (9). According to the literature [7], the weights $\lambda 1$, $\lambda 2$ and $\lambda 3$ of the loss function of the facial feature point fusion module are set to 0.1, 10 and 100, and 400 epochs are trained. In the training process of the expression transfer module, two metrics, namely structure similarity (SSIM) and peak signal to noise ratio (PSNR), are used to evaluate the generated facial expression images with the real ones. The SSIM reaches a maximum value of 0.805 and stabilizes after 100 epochs of training, and the PNSR also stabilizes at this time. Therefore, the loss function weights $\mu 1$ and $\mu 2$ of the expression transfer module are set to 1 and 100, and 120 epochs are trained.

Both modules use Adam [20] optimizer for parameter tuning. For the facial feature point fusion module, $\beta 1$ is 0.5, $\beta 2$ is 0.999, and Batch-Size is 8. For the expression transfer module, $\beta 1$ is 0.5, $\beta 2$ is 0.999, and Batch-Size is 8. The learning rate of the generator is 0.0002, and the learning rate of the discriminator is 0.0001.

C. RESULTS AND ANALYSIS

The proposed method is subjected to a series of experiments on the datasets CK+ with RaFD, and the effect of facial expression transfer is visualized. Figure 3 shows the effect of facial expression transfer on the CK+ dataset and Figure 4 shows the effect of facial expression transfer on the RaFD dataset, where the first row is the input source face image and the first column is the input target face image, target face images are randomly selected 2 from the training sets and 4 from the test sets. From figure 3 and figure 4, it can be intuitively seen that the faces generated by the model not only maintain the source expression and target face identity features, but also have natural and realistic expression.

The difference between the images generated by the facial expression transfer model and the real images is quantified using two metrics: structure similarity (SSIM) and peak signal to noise ratio (PSNR). The larger the PNSR value, the better the quality of the generated image. The proposed method is compared with CGAN, G2-GAN, and Pix2Pix models. Table 3 shows the comparison of evaluation metrics under the CK+ dataset and Table 4 shows the comparison of evaluation metrics under the RaFD dataset.

TABLE 3. Comparison of evaluation metrics under the CK+.

Model	SSIM	PSNR (dB)
CGAN	0.692	22.244
G2-GAN	0.767	24.420
Pix2pix	0.784	24.687
The proposed model	0.805	25.610

TABLE 4. Comparison of evaluation metrics under the RaFD.

Model	SSIM	PSNR (dB)
CGAN	0.554	16.412
G2-GAN	0.829	21.745
Pix2pix	0.765	19.230
The proposed model	0.801	19.338

The experimental results show that the proposed facial expression transfer model based on conditional generative adversarial network has an SSIM of 0.801 and PSNR of 19.338 dB on the RaFD dataset, both of which are better than the cGAN model and the Pix2pix model, but are relatively lower when compared to the G2-GAN. However, the SSIM of the model on the CK+ dataset is 0.805 and the PSNR is 25.610 dB, both metrics are significantly better than the other three models, especially than the cGAN. it can also be seen from Table 3 and Table 4 that the metrics on the CK+ dataset are better than those on the RaFD dataset for the same model, which is because the face images on the CK+ dataset are grayscale map, and the transfer of facial expression on grayscale images is easier than color images.

VI. CONCLUSION

The proposed facial expression transfer model based on conditional generative adversarial network can generate a highly realistic face image with source facial expression and target facial identity features to any given source face image and target face image. The model is designed with two closely related modules: the facial feature point fusion module and the expression transfer module, and also introduces the adversarial loss function, the cyclic consistency loss function and the pixel-level loss function, which can ensure that the synthesized face image keep the identity features of the target face well and make the facial expression more realistic and natural.

Further research will be conducted to incorporate and optimize facial attribute modification to allow the generation of face images with diverse styles by inputting relevant face attribute information.

REFERENCES

[1] X. X. Wang, F. F. Li, and Q. Chen, "Facial expression generation method based on enhanced conditional generative adversarial networks," *J. Chin. Comput. Syst.*, vol. 41, no. 9, pp. 1987–1992, Sep. 2020.

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, 2014, pp. 2672–2680.

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

[4] J. B. Chen, R. Liu, M. Liu, and Y. Feng, "Facial expression transfer model based on conditional generative adversarial network," *J. Comput. Eng.*, vol. 46, no. 4, pp. 228–235, Apr. 2020.

[5] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, Mar. 2010.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, 2015, pp. 234–241.

[7] F. Y. Dai, J. Chi, M. G. Ren, and Q. D. Zhang, "Face image synthesis driven by geometric feature and attribute label," *J. Comput. Sci.*, vol. 49, no. 10, pp. 214–223, Oct. 2022.

[8] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2Face: Real-time face capture and reenactment of RGB videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2387–2395.

[9] L. Tran and X. Liu, "Nonlinear 3D face morphable model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7346–7355.

[10] Z. Yu, J. Chi, Y. Ye, and F. Dai, "Detailed features-preserving 3D facial expression transfer," *J. Comput.-Aided Design Comput. Graph.*, vol. 33, no. 2, pp. 186–198, Feb. 2021.
[11] P. Isola, J. Y. Zhu, T. Zhou, and A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 1125–1134.
[12] L. Song, Z. Lu, R. He, Z. Sun, and T. Tan, "Geometry guided adversarial facial expression synthesis," in *Proc. 26th ACM Int. Conf. Multimedia*, New York, NY, USA, Oct. 2018, pp. 627–635.
[13] H. Ding, K. Sricharan, and R. Chellappa, "ExprGAN: Facial expression editing with controllable expression intensity," in *Proc. AAAI*, New Orleans, LA, USA, 2018, pp. 31–32.
[14] F. Qiao, N. Yao, Z. Jiao, Z. Li, H. Chen, and H. Wang, "Emotional facial expression transfer from a single image via generative adversarial nets: Emotional facial expression transfer from a single image via generative adversarial nets," *J. Comput. Animation Virtual Worlds*, vol. 29, nos. 3–4, p. e1819, May 2018.
[15] J. Liu and Z. Q. Jin, "Facial expression transfer method based on deep learning," *J. Comput. Sci.*, vol. 46, no. 6A, pp. 250–253. Jun. 2019.
[16] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.
[17] C. Aaron, C. Faruk, A. Martin, D. Vincent, and I. Gulrajani, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 5769–5779.
[18] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.-Workshops*, San Francisco, CA, USA, Jun. 2010, pp. 94–101.
[19] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the Radboud faces database," *Cognition Emotion*, vol. 24, no. 8, pp. 1377–1388, Dec. 2010.
[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

YANG FAN received the B.S. degree from the Sichuan University of Science and Engineering, Sichuan, China, in 2021, where he is currently pursuing the degree with the School of Automation and Electric Information. His current research interests include image processing and deep learning.

XINGGUO JIANG received the M.S. degree from Chongqing University, Chongqing, China, in 2003, and the Ph.D. degree from the Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, China, in 2007. He is currently an Associate Professor with the School of Automation and Electric Information, Sichuan University of Science and Engineering, Sichuan, China. Prior to that, he was an Associate Professor with the School of Information and Communication, Guilin University of Electronic Technology, Guangxi, China. His current research interests include image processing, intelligent information processing, and deep learning.

SHUXING LAN received the B.S. degree from the Sichuan University of Science and Engineering, Sichuan, China, in 2021, where he is currently pursuing the degree with the School of Automation and Electric Information. His current research interests include image processing and deep learning.

JIANGHAI LAN was born in Sichuan, China, in 1995. He received the B.S. degree from the Sichuan University of Science and Engineering, in 2021, where he is currently pursuing the degree with the School of Automation and Information Engineering. His current research interests include image processing and deep learning.