

Research Article

Improved Facial Expression Recognition Method Based on GAN

Junhuan Wang 

School of Computer Science, Guangdong University of Education, Guangzhou 510303, Guangdong, China

Correspondence should be addressed to Junhuan Wang; wangjunhuan06@163.com

Received 10 November 2021; Revised 26 November 2021; Accepted 30 November 2021; Published 21 December 2021

Academic Editor: Le Sun

Copyright © 2021 Junhuan Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recognizing facial expressions accurately and effectively is of great significance to medical and other fields. Aiming at problem of low accuracy of face recognition in traditional methods, an improved facial expression recognition method is proposed. The proposed method conducts continuous confrontation training between the discriminator structure and the generator structure of the generative adversarial networks (GANs) to ensure enhanced extraction of image features of detected data set. Then, the high-accuracy recognition of facial expressions is realized. To reduce the amount of calculation, GAN generator is improved based on idea of residual network. The image is first reduced in dimension and then processed to ensure the high accuracy of the recognition method and improve real-time performance. Experimental part of the thesis uses JAFEE dataset, CK+ dataset, and FER2013 dataset for simulation verification. The proposed recognition method shows obvious advantages in data sets of different sizes. The average recognition accuracy rates are 96.6%, 95.6%, and 72.8%, respectively. It proves that the method proposed has a generalization ability.

1. Introduction

Recognizing facial expressions can provide a more comprehensive understanding of people's inner world [1]. It has many applications in medicine, transportation, culture, and education [2–5]. Therefore, the recognition and analysis of facial expressions have important research significance and value.

At present, certain researchers have carried out research on image expression recognition, the purpose of which is to accurately classify and recognize seven basic emotional expressions in facial images [6, 7] including anger, disgust, fear, happy, sad, surprise, and neutrality.

Traditional facial expression feature extraction uses mathematical methods to calculate and process facial expression images. Mainly, it can be divided into two situations for processing static images and dynamic images. Statistical methods, Gabor wavelet, and local binary method belong to the feature extraction of static images [8, 9]. Geometric method, optical flow method, and model method belong to the feature extraction of dynamic images [10–12]. However, there are diversity and complexity in image acquisition. Traditional facial expression recognition methods face the

problem of nonlinear uncertainty of sample data. The features selected in facial expression feature extraction have no good representation ability [13] and need to be extracted manually according to people's experience. These problems will have a great impact on the recognition accuracy of the model, resulting in the poor generalization ability.

It should be noted that for facial expression recognition research, its essence is to optimize and analyze massive data [14, 15]. Benefiting from the development of artificial intelligence and big data technology, the deep network model can effectively extract effective image features from massive multidimensional image data through continuous iterative learning of multilayer network. And based on its strong learning ability, compared with the traditional facial expression recognition method, it can classify the facial expression more accurately and quickly [16–19]. Reference [20] analyzed the facial expression information of time series based on the part-based hierarchical bidirectional recurrent neural network and extracts facial temporal features in dataset. It can comprehensively analyze facial expressions. Reference [21] proposed a method based on fusion of deep belief network (DBN) and local features. This method extracts eyebrows, eyes, and mouth with rich expression

information as local expression images. It also combines the Log-Gabor feature with texture information and the second-order histogram with the gradient direction feature of shape information to realize facial expression recognition. Reference [22] combined spatio-temporal features and used deep residual networks to extract feature. Reference [23] used three channels to extract feature of expression images, respectively. Then, extracted features are connected and sent to the next layer for processing.

Considering the previous work, the network designed in this paper mainly has the following innovations:

- (1) Through continuous confrontation training of generator and discriminator in GAN, the deep extraction of the features of processed expression dataset is realized thereby improving performance of corresponding facial expression classification and recognition.
- (2) The performance and speed of network are improved. The idea of adding residual network to the generator network improves the operation efficiency while ensuring accuracy.

In addition, the second section introduces related theoretical methods. The third section introduces the improved network and explains the structure of the discriminator and generator in the network. The fourth section introduces the experimental results and analysis based on different datasets. The fifth section is the conclusion.

2. Related Methods and Theories

2.1. General Steps. Face recognition technology includes four steps. They are face detection, face alignment, face representation, and face matching, as shown in Figure 1. Face detection module is used to detect position of face in input image. The face alignment module automatically locates key points around the face according to the input such as eyebrows, eyes, corner points of mouth, nose tip, and contour points. Face characterization is to locate a face picture from the above two steps, extract from it, or convert it into a feature vector. In the face matching part, the extracted feature vector will be used to compare with these in database. Based on the similarity between the two, it is possible to determine whether they belong to the same person in the database.

The image needs to be preprocessed to improve accuracy of judgment [24]. The advantages of traditional face key point detection algorithms are clear architecture and easy to understand. However, the operation efficiency is not high, and it is not suitable for processing a large number of images.

For face characterization processing and analysis, the data feature vector often contains information such as the position of eyebrows, nose, and eyes and even additional information such as contour and shape. The more classic methods include HOG method, Haar wavelet method, and eigenface method. However, traditional methods are used to extract the front of a human face, but the effect is not good enough for the side [25].

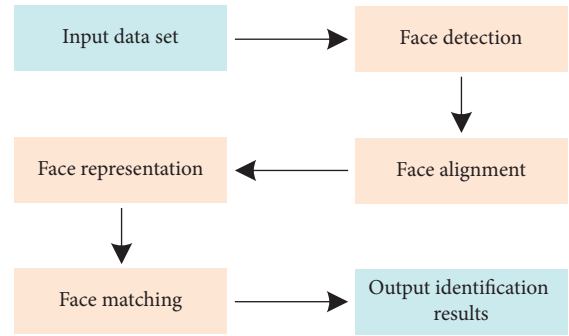


FIGURE 1: The step of face recognition.

Face matching generally compares the extracted face feature vector with these in database. If the distance of feature vector is close, identity information is output. If there is no match for all faces in database, output cannot be recognized.

2.2. Convolutional Neural Network (CNN). CNN is good at processing images [26]. Traditional face recognition methods show poor results when facing complex scenes. CNN-based deep learning methods can automatically extract features based on a large amount of image data and perform well in complex scenes.

The CNN model is essentially a deep feedforward model, which updates parameters through backpropagation. To obtain better results, it generally needs to design the cores of convolutional layer and pooling layer. And they will be continuously combined to obtain better image characteristics.

2.3. Generative Adversarial Network. Due to its multilayer network structure, the convolutional neural network also has too many problems in its network parameter settings, which makes the CNN face recognition training process very fragile [27]. For face recognition research and analysis, subtle changes in the CNN structure or a little adjustment of parameters will lead to deviations in the recognition results.

As a deep learning model that is widely used in current image analysis, GAN can solve the problem of instability in the training process through adversarial learning methods.

A typical GAN consists of two part, namely, generator G and discriminator D . During training, these two subnetworks play a game, as shown in Figure 2.

First, the generated image and real image are input into discriminator at the same time, and the discriminator is trained. As the training process progresses, the pictures generated by the generator become more and more realistic, and the classification ability of the discriminator is gradually improved. Finally, the training process reaches a state of convergence. The discriminator cannot identify the true and false of the input image, and the image generated is also the same as the real image; that is, the Nash equilibrium state is reached. The training process of the entire game can be described in the following value function $V(D, G)$:

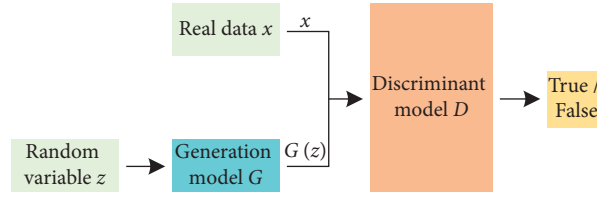


FIGURE 2: Schematic diagram of generative adversarial networks.

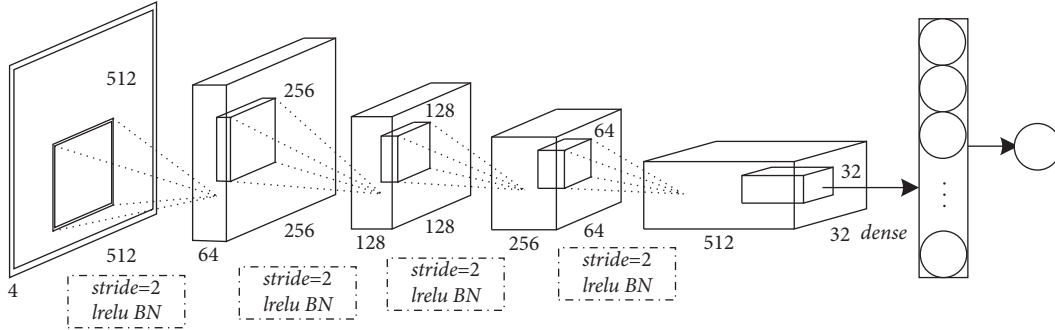


FIGURE 3: Discriminator network structure.

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

where $E_{x \sim p_{data}(x)}$ and $E_{z \sim p_z(z)}$ are the expected functions; x is the real image; z is the image of input generator. G converts the variable z into the probability $G(z)$ that the image generated by the converter is a real image. The variable z is basically a sample from the distribution p_z . The ideal distribution p_z should converge to the data distribution p_{data} . Practice has proved that in the generator, maximizing the logarithm $\log(D(G(z)))$ is better than minimizing the logarithm $\log(1 - D(G(z)))$.

Since the GAN network has two models, the loss of discriminator is as follows:

$$\min_D V(D, G) = E_{x \sim p_{data}(x)} \{\log D(x)\} + E_{x \sim p_{G(x)}} \{\log\{1 - D[G(z)]\}\}. \quad (2)$$

When training loss function of generator, default discriminator has the best ability. The $E_{x \sim p_{data}(x)} \{\log D(x)\}$ part is a constant, so the loss of generator is as follows:

$$\min_G V(D, G) = E_{x \sim p_{G(z)}} \{\log 1 - D[G(z)]\}. \quad (3)$$

3. Method

3.1. Discriminator Network. For the discriminator network in the GAN, this paper uses the VGG-16 network as backbone network structure [28], and the network structure is shown in Figure 3. Using $\{x_i, y_i\}_{i=1}^N$ or $\{x_i, G(x_i)\}_{i=1}^N$ ($x_i \in X, y_i \in Y$) as input, $x \in R^{512 \times 512 \times 3}$ and $y \in R^{512 \times 512 \times 1}$. When the input is $\{x_i, y_i\}$, correct output of discriminator is 1 and the correct output of input $\{x_i, G(x_i)\}$ discriminator is 0. Leaky-ReLU is used as a nonlinear

activation function in each convolutional layer in the discriminator.

First, two convolution and pooling operations are performed on the image, and each operation includes two convolutions and one maximum pooling. Then, three convolution and pooling operations are performed, and each operation includes three convolution operations and a maximum pooling operation. Finally, there are three fully connected layers and one Softmax layer. Similar to traditional generative confrontation network, the discriminator mainly judges the authenticity of input discriminator image. Input image has same size and dimension as the generated image, and both are $3 \times 48 \times 48$. The adversarial loss E_{adv} is defined as follows:

$$E_{adv}(\theta_f, \theta_h, \theta_d) = \sum_{i=1,2,\dots,N} (1 - L_d(G_d(x_i; \theta_d))) + \sum_{i=1,2,\dots,N} L_d(G_d(G_h(G_f(x_i; \theta_f); \theta_h); \theta_d)), \quad (4)$$

where x_i is the real image, G_f is the feature extractor, θ_f is the parameter of feature extractor, G_h is a feature synthesizer, θ_h is the parameter of feature synthesizer, G_d is the discriminator, θ_d is the parameter of discriminator, and L_d is the loss calculation function of discriminator. Then, the total loss function E_{total} is as follows:

$$E_{total}(\theta_f, \theta_h, \theta_c, \theta_d) = E_c(\theta_f, \theta_c) + E_{adv}(\theta_f, \theta_h, \theta_d). \quad (5)$$

3.2. Generator Network. The generator network in GAN uses $x \in R^{w \times h \times c}$ as the input image, where $w = h = 512, c = 3$. The network structure is shown in Figure 4. Some previous segmentation methods use encoder-decoder [29]. This structure first down-samples and then gradually up-samples.

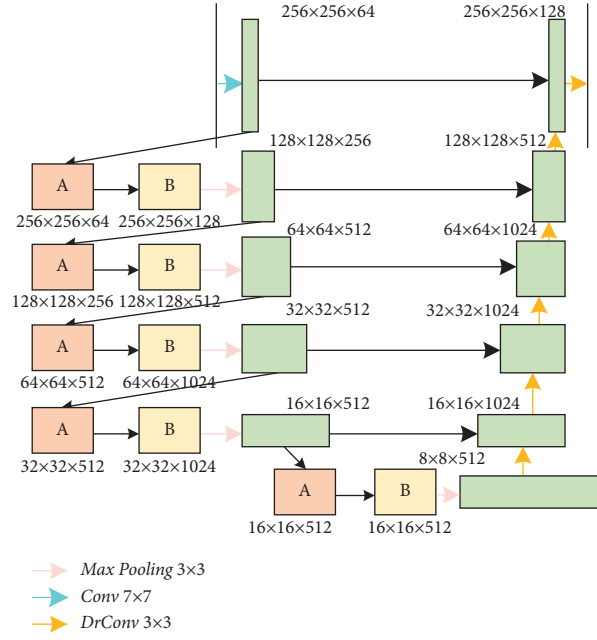


FIGURE 4: Generator network structure.

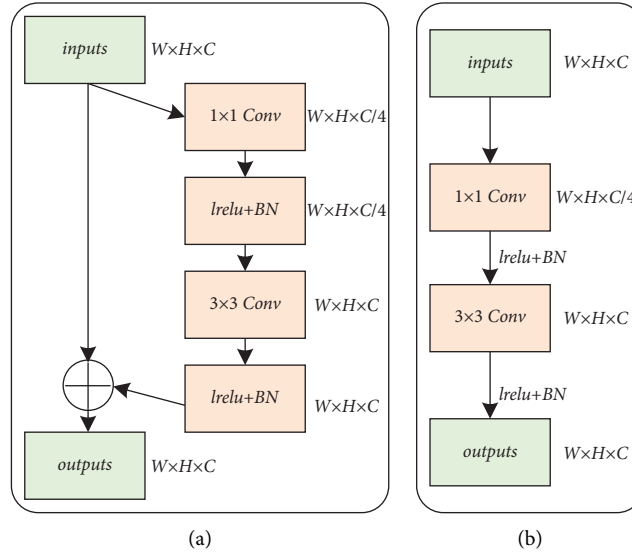


FIGURE 5: Different variant of convolutional.

This paper uses a U-shaped structure for the generator. The feature extractor is used to extract the feature of input image. Image input resolution is $3 \times 48 \times 48$, and the backbone network uses ResNet-18 [30]. Unlike the traditional generative confrontation network, the generator input is not random noise but a facial expression image. First, the feature extractor performs a 3×3 convolution operation on the input image $x_i (i = 1, 2, \dots, n)$ with a step size of 1. Then, there are batch normalization and ReLU. Second, convolution operation of 4 modules is performed, respectively. Then, the average pooling operation is performed after convolution, and window size is 2×2 . Dropout is used after the average pooling operation.

Finally, the extracted features are input to two fully connected layers and one Softmax layer. The 512-dimensional feature vector is classified into 7 types of facial expressions, and the facial expression recognition results are obtained. The classification loss E_c of classifier is defined as

$$E_c(\theta_f, \theta_c) = \sum_{i=1,2,\dots,N} L_c(G_c(G_f(x_i; \theta_f); \theta_c), y_i), \quad (6)$$

where x_i is the original input image, θ_f is the parameter of feature extractor, G_f is the feature extractor, θ_c is the parameter of classifier, G_c is the classifier, y_i is the real label, and L_c is the classification loss.

At the same time, this paper adds a residual module to generator, and it is shown in Figure 5(a). The structure of the forward propagation convolution unit is shown in Figure 5(b).

Through the confrontation training between the generator and the discriminator, the feature extractor's ability to extract features and the discriminator's recognition ability are improved. The feature synthesizer is a symmetrical structure to the feature extractor and is mainly composed of a convolutional layer and an upsampling layer. After continuous convolution and upsampling operations, the generated output image is restored to the original size.

4. Experimental Results and Analysis

The experiment uses TensorFlow framework to implement network model training on simulation dataset. To ensure quality of the experiment, *Python* is used as the programming language. And NVIDIA CUDA 9.0 is used for GPU accelerated computing. The specific system development environment of the face recognition simulation experiment is shown in Table 1.

4.1. Parameter Setting. When the face recognition network is trained, the optimization method uses SGD, momentum parameter is set to 0.9, and weight decay rate is 10^{-4} . Learning rate is a reduction strategy of multiplying the initial $lr = 3 \times 10^{-3}$ by $(1 - \text{current_iter}/\text{max_iter})^{\text{power}}$ where $\text{power} = 0.9$, current_iter is the current number of iterations, and max_iter is the maximum number of iterations in training process. For the discriminant network, the Adam optimization method is used, $\text{betas} \in (0.9, 0.99)$, and the initial $lr = 1 \times 10^{-4}$. The learning rate reduction strategy is the same as the method of training the segmentation network. Taking into account the GPU memory limitation, the image size in the experiment is set to 348×348 pixels.

4.2. Evaluation Index. To measure performance of identification our method, an objective and fair evaluation index should be used. Accuracy (AC), precision (P), and recall (R) are commonly used indicators in big data image classification research, which can be used to analyze performance of face recognition results. The calculation formulas are shown in formulas (7)–(9).

P represents how many of the samples that the model predicts to be positive are true categories. R is expressed as how many of the model's predicted categories are positive examples in the samples where the true category is positive.

$$AC = \frac{TP + TN}{TP + FN + TN + FP}, \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (9)$$

TABLE 1: System development environment.

Hardware environment	Software environment
CPU: AMD 7 4800H 2.9 GHz	Operating system: Windows 10
Graphics card: NVIDIA CUDA 9.0	Development tool: TensorFlow
Memory card: 16 G	Development language: <i>Python</i>

TABLE 2: Confusion matrix.

Real situation	Forecast result	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

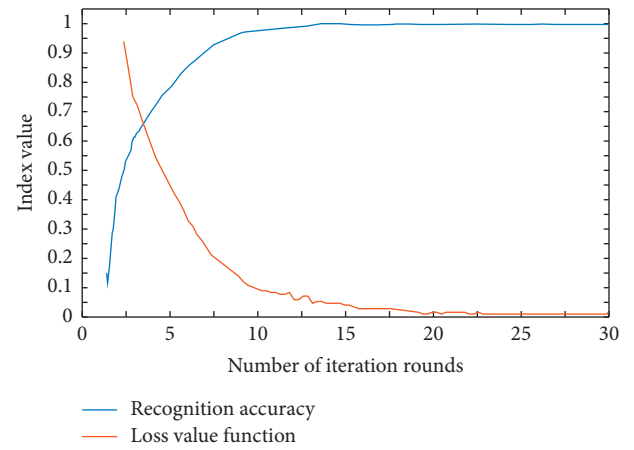


FIGURE 6: Convergence of the proposed method under multiple iterations.

For classification problems, the combination of the model prediction result and the true category of the sample can be divided into true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

The precision and recall rate can be represented by a confusion matrix, as shown in Table 2.

At the same time, the loss value function c is used to evaluate the model and to measure the quality of the training performance of the GAN model. The appropriate number of iterations is determined in the process of training discrimination. In this paper, cross-entropy loss is used to express probability of predicting which type of input sample belongs to, and its expression is as follows:

$$c = -\frac{1}{n} \sum [y \ln a + (1 - y) \ln (1 - a)], \quad (10)$$

where y is the true classification value, a is the predicted value, and c represents the loss value.

4.3. Training Process. We analyze recognition and classification performance of the GAN model for face collection data and explore the convergence of collection data training

TABLE 3: Identification performance based on JAFFE dataset.

	Average recognition accuracy (%)	Running time (s)
The proposed method	96.6	54.23
Reference [21]	95.7	54.53
Reference [22]	93.5	55.32
Reference [23]	94.3	56.32

TABLE 4: Identification performance based on CK + dataset.

	Average recognition accuracy (%)	Running time (s)
The proposed method	95.6	84.23
Reference [21]	91.7	89.53
Reference [22]	90.5	88.32
Reference [23]	90.3	84.32

process. Figure 6 shows the convergence performance of the training process for the expression dataset.

In the 10th iteration, recognition accuracy of training set samples has reached 95%. At the end of 15 iterations, the accuracy is approximately close to 100%. At the same time, through the numerical analysis of the loss function of each iteration, it can be known that the training set has been quickly and effectively attenuated before the 10th iteration. At the 18th iteration, the loss function value is close to 0. In summary, the improved expression recognition method of GAN has good convergence performance.

4.4. Simulation Analysis of General Experimental Dataset.

The experimental simulation analysis is carried out using the methods proposed in references [21–23] and this paper. To verify generalization performance of the proposed recognition method on data sets of different sizes, the small, medium, and large experimental simulation data sets are selected as JAFFE dataset, CK + dataset, and FER2013 dataset in turn.

The JAFFE dataset was created by the Michael Lyons team. The image data collected in this dataset contain the expressions of 10 Japanese female participants, with a total of 213 facial images. The JAFFE dataset contains 7 types of basic expressions: anger, happy, sad, surprise, fear, disgust, and neutral.

The CK + dataset comes from the Patrick Lucey team’s expansion in the Cohn–Kanade dataset. The CK + dataset collected 123 facial expression images of different people, a total of 593 expression sequences and 951 image samples. The image pixel size is $3 \times 48 \times 48$.

The FER 2013 dataset comes from the Kaggle competition and consists of 35886 facial expression pictures. There are 28708 test sets, 3589 public verification sets, and 3589 private verification sets. Each image is composed of 48×48 grayscale images.

4.4.1. JAFFE Dataset Experiment. This paper chooses JAFFE dataset as a small dataset to simulate and verify performance of the proposed GAN facial expression recognition method. Table 3 shows the stability results based on JAFFE dataset under different methods.

It can be seen from Table 3 that in terms of facial expression recognition for JAFFE dataset, accuracy of our method is 96.6%. It is 0.9%, 2.1%, and 2.3% higher than references [21–23]. The proposed method has no obvious advantage over the comparative method in terms of simulation runtime. Therefore, we believe that when performing expression recognition on small data sets, the method proposed in this article can be selected for efficient discrimination.

4.4.2. CK + Dataset Experiment. The CK + dataset is used as a medium-sized data set for facial expression recognition in this article, and different methods are also used for comparative analysis with our method. The face recognition performance of the CK + dataset under different methods is shown in Table 4.

It can be seen from Table 4 that our method has the accuracy of 95.6% for CK + dataset, which is 5.3% higher than that in reference [23]. The PCNN used in reference [23] has more network layers. There is the problem of the disappearance of the network gradient during training, which causes a large gap in the recognition accuracy compared with our method. The simulation time of the identification method in this paper is 84.23 s, which is 5.3 s shorter than that in reference [21]. Compared with reference [23], the simulation time is relatively close, but reference [23] does not have an advantage in recognition accuracy. Therefore, it is proved that GAN has good accuracy and real-time performance for facial expression recognition of medium-sized volume data sets.

4.4.3. FER2013 Dataset Experiment. Table 5 shows the simulation analysis results of large datasets using different facial expression recognition and classification methods.

From Table 5, the accuracy of expression classification and recognition of all methods for the FER2013 dataset is both below 75%. This is because there are a certain number of error labels in the FER2013 dataset. All of this results in a lower accuracy of the recognition method. However, our method has the highest accuracy of 72.8%. The running time of our method is 134.23 s, which is shortened by more than 10 s

TABLE 5: Identification performance based on FER2013 dataset.

	Average recognition accuracy (%)	Running time (s)
The proposed method	72.8	134.23
Reference [21]	67.8	145.53
Reference [22]	69.5	144.32
Reference [23]	65.3	146.32

Anger	0.65	0.00	0.10	0.03	0.11	0.01	0.09
Hate	0.23	0.62	0.02	0.02	0.11	0.00	0.00
Fear	0.09	0.00	0.57	0.02	0.17	0.07	0.08
Happy	0.02	0.00	0.01	0.88	0.01	0.02	0.06
Sadness	0.10	0.00	0.11	0.03	0.58	0.02	0.16
Surprise	0.02	0.00	0.06	0.03	0.02	0.85	0.02
Neutral	0.06	0.00	0.04	0.07	0.15	0.01	0.67
	Anger	Hate	Fear	Happy	Sadness	Surprise	Neutral

FIGURE 7: Schematic diagram of the confusion matrix of FER2013 dataset.

compared with references [21–23]. Therefore, the improved expression recognition method of GAN proposed in this paper can also be used for large-scale data set analysis.

To further illustrate recognition performance, a confusion matrix is used to display and illustrate the recognition results obtained by our method, as shown in Figure 7. The accuracy of the method for the recognition of anger, disgust, fear, happiness, sadness, surprise, and neutral expressions is 65%, 62%, 57%, 88%, 58%, 85%, and 67%, respectively.

Figure 7 shows that the method performs well in identifying “happy” and “surprised,” with accuracy rates reaching 88% and 85%, respectively. In addition, it can be noticed that the ability to recognize “fear” facial expressions of generating confrontation network is low, with an accuracy rate of 57%. This is because the labeling in the FER 2013 dataset is not good.

In summary, compared with other methods, our method has higher accuracy and operating efficiency for different volume data sets. It shows that our method has excellent generalization ability.

5. Conclusion

This paper proposes a facial expression recognition method based on GAN. This method is based on continuous confrontation training between generator structure and discriminator structure in GAN, which realizes the accurate extraction of data set features and ensures the accurate recognition of facial expressions. By improving the

generator structure in GAN network, the residual network is combined with image processing technology. Thus, the amount of calculation for identifying the network model is reduced. Finally, based on general datasets of different sizes, our method is validated for the efficient performance of facial expression recognition. It is proved that our method has obvious advantages in recognition accuracy and processing speed.

In the future, we also plan to add an attention mechanism to the network to further improve accuracy and prune the network to improve efficiency and strive to achieve industrialization.

Data Availability

The data included in this paper are available without any restriction.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the “Software Engineering” (Key Subjects’ Construction Project) in Guangdong University of Education.

References

- [1] I. Gogic, M. Manhart, I. S. Pandzic, and J. Ahlberg, “Fast facial expression recognition using local binary features and shallow neural networks,” *The Visual Computer*, vol. 36, no. 1, pp. 97–112, 2020.
- [2] J. Lee, S. Kim, S. Kim, and K. Sohn, “Multi-modal recurrent attention networks for facial expression recognition,” *IEEE Transactions on Image Processing*, vol. 29, no. 1, pp. 6977–6991, 2020.
- [3] Y. Yan, Y. Huang, S. Chen, C. Shen, and H. Wang, “Joint deep learning of facial expression synthesis and recognition,” *IEEE Transactions on Multimedia*, vol. 22, no. 11, pp. 2792–2807, 2020.
- [4] A. R. Does, F. Barbosa, C. Queirós, I. P. Carvalho, and M. D. Griffiths, “Recognizing emotions through facial expressions: a l experimental study,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 20, pp. 1–13, 2020.
- [5] L. Yan, “Application of face expression recognition technology in skilled unsupervised course based on ultra-wide regression network,” *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 6, pp. 7167–7177, 2020.

- [6] G. Ali, A. Ali, F. Ali et al., "Artificial neural network based ensemble approach for multicultural facial expressions analysis," *IEEE Access*, vol. 8, no. 1, pp. 134950–134963, 2020.
- [7] Y. T. Su, M. M. Wang, J. Liu, Y.P. Fei, and X. He, "Micro expression recognition algorithm based on multi motion feature fusion," *Progress of laser and optoelectronics*, vol. 57, no. 14, pp. 353–358, 2020.
- [8] T. Chen, T. Gao, and X. Zhao, "Single sample description based on Gabor fusion," *IET Image Processing*, vol. 13, no. 14, pp. 2840–2849, 2020.
- [9] F. Kong, "Facial expression recognition method based on deep convolutional neural network combined with improved LBP features," *Personal and Ubiquitous Computing*, vol. 23, no. 3–4, pp. 531–539, 2019.
- [10] M. H. Kabir, M. S. Salekin, M. Z. Uddin, and M. Abdullah-Al-Wadud, "Facial expression recognition from depth video with patterns of oriented motion flow," *IEEE ACCESS*, vol. 5, no. 1, pp. 8880–8889, 2017.
- [11] X. Jia, X. Ben, H. Yuan, K. Kpalma, and W. Meng, "Macro-to-micro transformation model for micro-expression recognition," *Journal of computational science*, vol. 25, no. 1, pp. 289–297, 2018.
- [12] E. Zangeneh and A. Moradi, "Facial expression recognition by using differential geometric features," *The Imaging Science Journal*, vol. 66, no. 8, pp. 463–470, 2018.
- [13] N. Y. Zheng, H. Zhang, B. Y. Song, W. Liu, Y. Li, and A. M. Dobaie, "Facial expression recognition via learning deep sparse autoencoders," *Neurocomputing*, vol. 273, no. 1, pp. 643–649, 2018.
- [14] J. Guo, K. Shan, H. Wu, R. Bie, W. You, and D. Lu, "Research on facial expression recognition technology based on convolutional-neural-network structure," *International Journal of Software Innovation*, vol. 6, no. 4, pp. 103–116, 2018.
- [15] J. H. Lu, S. M. Zhang, and J. L. Zhao, "Facial expression recognition based on deep learning," *Computer application research*, vol. 37, no. 04, pp. 966–972, 2020.
- [16] Y. He and S. X. Chen, "Person-independent facial expression recognition based on improved local binary pattern and higher-order singular value decomposition," *IEEE ACCESS*, vol. 8, no. 1, pp. 190184–190193, 2020.
- [17] Z. Zhang, C. Lai, H. Liu, and Y.-F. Li, "Infrared facial expression recognition via Gaussian-based label distribution learning in the dark illumination environment for human emotion detection," *Neurocomputing*, vol. 409, no. 1, pp. 341–350, 2020.
- [18] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with Convolutional Neural Networks: c," *Pattern Recognition*, vol. 61, no. 1, pp. 610–628, 2017.
- [19] X. L. Peng, H. X. Zhang, and Q. Y. Hu, "Face/facial paralysis expression recognition method based on multi task deep convolution neural network," *Journal of Northwest University*, vol. 49, no. 2, pp. 187–192, 2019.
- [20] K. Zhang, Y. Huang, Y. Du, and L. Wang, "Facial expression recognition based on deep evolutionary spatial-temporal networks," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4193–4203, 2017.
- [21] L. L. Wang, J. H. Liu, and X. M. Fu, "Facial expression recognition based on local features and deep belief networks," *Progress in laser and optoelectronics*, vol. 55, no. 1, pp. 204–212, 2018.
- [22] Y. H. Ma, R. J. Zhang, C. Wu, and J.S. Qu, "Expression recognition of image sequence based on deep residual network and LST," *Journal of Chongqing University of Posts and Telecommunications (NATURAL SCIENCE EDITION)*, vol. 32, no. 5, pp. 874–883, 2020.
- [23] L. L. Xu, S. M. Zhang, and J. L. Zhao, "Expression recognition algorithm based on parallel convolutional neural network," *Chinese Journal of image and graphics*, vol. 24, no. 2, pp. 227–236, 2019.
- [24] W. Zhang and Y. T. Qian, "Streamlined face landmark detection network with global constraint," *Journal of Signal Processing*, vol. 35, no. 3, pp. 507–515, 2019.
- [25] M. Annalakshmi, S. M. M. Roomi, and A. S. Naveedh, "A hybrid technique for gender classification with SLBP and HOG features," *Cluster Computing*, vol. 22, no. 1, pp. 11–20, 2019.
- [26] T. Nishime, S. Endo, K. Yamada, N. Toma, and Y. Akamine, "Feature acquisition from facial expression image using convolutional neural networks," *Journal of Robotics, Networking and Artificial Life*, vol. 3, no. 1, pp. 9–12, 2016.
- [27] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for h image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5046–5063, 2018.
- [28] D. Chun, T. S. Kim, K. Lee, and H.-J. Lee, "Compressed video restoration using a generative adversarial network for subjective quality enhancement," *IEIE Transactions on Smart Processing & Computing*, vol. 9, no. 1, pp. 1–6, 2020.
- [29] C. Tao, H. Wang, J. Qi, and H. Li, "Semisupervised v generative adversarial networks for h image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, no. 1, pp. 914–927, 2020.
- [30] L. Ma, R. Shuai, X. Ran, W. Liu, and C. Ye, "Combining DC-GAN with ResNet for blood cell image classification," *Medical, & Biological Engineering & Computing*, vol. 58, no. 6, pp. 1251–1264, 2020.