Hindawi
Journal of Robotics
Volume 2021, Article ID 9326695, 9 pages
https://doi.org/10.1155/2021/9326695

Research Article

Expression Recognition Method Using Improved VGG16 Network Model in Robot Interaction

Shengbin Wu

School of Information Engineering, Changsha Medical University, Changsha, Hunan 410219, China

Correspondence should be addressed to Shengbin Wu; shengbinwu123456@163.com

Received 12 November 2021; Accepted 4 December 2021; Published 20 December 2021

Academic Editor: Shan Zhong

Aiming at the problems of poor representation ability and less feature data when traditional expression recognition methods are applied to intelligent applications, an expression recognition method based on improved VGG16 network is proposed. Firstly, the VGG16 network is improved by using large convolution kernel instead of small convolution kernel and reducing some fully connected layers to reduce the complexity and parameters of the model. Then, the high-dimensional abstract feature data output by the improved VGG16 is input into the convolution neural network (CNN) for training, so as to output the expression types with high accuracy. Finally, the expression recognition method combined with the improved VGG16 and CNN model is applied to the human-computer interaction of the NAO robot. The robot makes different interactive actions according to different expressions. The experimental results based on CK + dataset show that the improved VGG16 network has strong supervised learning ability. It can extract features well for different expression types, and its overall recognition accuracy is close to 90%. Through multiple tests, the interactive results show that the robot can stably recognize emotions and make corresponding action interactions.

1. Introduction

Different facial expressions can reflect people's emotional and psychological changes in different situations. Therefore, expression recognition has very important research significance and practical application value for the study of human behavior and psychological activities. In recent years, with the rapid development of computer vision, the rise of deep learning, the improvement of machine learning, and other related theoretical systems, facial expression, as a bridge of human-computer interaction, has attracted the attention of researchers at home and abroad. Facial expression recognition (FER) technology has developed rapidly [1, 2]. Different methods are used to extract facial expression feature information, and different methods are used to recognize and classify facial expressions, so as to make better use of emotional information and apply expression recognition technology to actual production, work, and life. With the development of deep learning, FER has gradually become one of the hot technologies in the field of human-computer interaction technology. FER has been widely used in human-computer interaction, business, medical treatment, entertainment, psychology, fatigue driving, and other fields [3, 4].

With the prevalence of artificial intelligence and pattern recognition, FER has encountered many practical problems in the application process, such as the sharp increase of data leads to the reduction of recognition effect. Therefore, researchers need to constantly improve algorithm to improve its recognition rate and robustness. There are two traditional FER methods: geometry-based method and whole-based recognition method [5]. These two methods rely on the advantages and disadvantages of the previous manual feature extraction, are subject to many interference factors in feature extraction, and are not realistic for massive image data processing [6]. With the continuous upgrading of computer technology, the improvement of computing power and communication ability has promoted the development of machine learning, especially, deep learning has made outstanding contributions to facial expression recognition [7, 8]. However, at present, most expression recognition methods based on deep learning still have great room for improvement in training samples and recognition

efficiency. Therefore, this paper proposes an expression recognition method based on the improved VGG16 network model.

2. Related Works

Researchers at home and abroad have carried out research studies on facial expression recognition. Traditional expression recognition methods mainly include geometry-based and overall FER methods. Its main advantage is that it is simple and easy to implement, but there is still room for improvement in recognition accuracy and efficiency [9]. For example, Madzin et al. [10] proposed a FER method based on geometric skin color to accurately detect faces in real-world images, which can adapt to image face recognition under different indoor and outdoor lighting environments and complex background conditions. However, the extraction and recognition of facial expression features cannot be carried out at the same time, which greatly limits the recognition efficiency.

Different from the traditional recognition methods, deep learning breaks the fixed pattern of pattern recognition after traditional feature extraction, and can carry out feature extraction and expression classification at the same time. Moreover, the feature extraction of deep learning iteratively updates the weights through back-propagation algorithm and error optimization, so it can extract key points and features unexpected to human beings [11, 12]. Chen et al. [13] proposed a simplified face clipping and rotation strategy combined with the image recognition method of convolutional neural network (CNN), which ensures the richness of data while extracting facial features and considers the accuracy and richness of expression recognition, but the training time for some complex expressions is long. Xu et al. [14] proposed a microexpression recognition method based on the Wasserstein generative adversarial network, established facial expression recognition network and facial identity recognition network, and improved the accuracy and robustness of facial expression recognition by suppressing intraclass changes. Lee et al. [15] converted the expression image into a local binary pattern (LBP) feature map and then used the LBP feature map as the input of CNN for training, which had achieved good results. However, it will lead to low accuracy and insufficient robustness in unknown environment. Bunyak et al. [16] proposed an automatic screen expression recognition system based on deep learning. On the basis of extracting the basic features of facial expressions, the original facial images are used to classify facial expressions and icon facial images to realize real-time recognition of facial expressions, but the overall accuracy needs to be improved. For facial expression recognition, Shao and Qian [17] proposed three facial recognition frameworks: shallow CNN, dual branch CNN, and pretrained CNN. The results show that they have good practicability and effectiveness. Unwala et al. [18] proposed a facial expression recognition system based on depth CNN. This paper constructs a dual channel CNN model and gathers the channel feature information into the full-connection layer for expression classification. The results show that it obtains high expression recognition accuracy, but it needs large training dataset and long training time in the early training process, so the recognition efficiency is not high.

In view of the low recognition accuracy of most existing expression recognition methods, this paper proposes an expression recognition method using the improved VGG16 network model, which improves the overall effect of expression recognition on the basis of taking into account the accuracy and recognition efficiency.

3. Proposed Method

3.1. Overall Network Structure. Deep learning perfectly solves the problem of automatic feature extraction and has a certain effect on face recognition and prediction. However, in the field of face expression recognition, the existing implementation effect does not reach the level required by the production environment [19]. With the continuous development of deep learning theory and structure research, this paper proposes an improved VGG16 model for facial expression recognition and makes in-depth research on expression feature extraction to effectively solve the problems of weak feature representation ability of machine learning extraction. The overall network structure is mainly composed of VGG16 and CNN. VGG16 is used for feature extraction, and CNN is used for classification and recognition. The specific structure is shown in Figure 1.

Compared with other deep learning methods for feature extraction, the proposed method uses the VGG16 network for feature extraction, which uses a large number of small-size convolution and pooling operations, so that VGG16 network can extract more implicit features, then uses CNN network for facial expression model training, and finally uses Softmax classifier for classification and recognition.

3.2. Data Preprocessing. In order to prevent the overfitting phenomenon of the network model, the data are randomly turned horizontally and cut or rotated at an angle during the training process. This method is called data enhancement. The data enhancement methods of horizontal, clipping, and angular rotation are adopted to enable the network model to learn the expression features from multiple angles, so as to enhance the robustness of the network model [20]. Data reduction can expand the number of images in the dataset, so that the neural network can learn more comprehensive feature information. Therefore, the dataset is usually randomly reduced in the experiment.

The size of the facial expression image after the facial expression data set CK + alignment is moderate, so the image is not enlarged in the experiment to retain the original information of the image to the greatest extent. When training, the original 100×100 images are randomly reduced to 90×90 size sub-image, then the image is randomly mirrored, and the processed image is sent to the network model for training. At the same time, in the test stage of the experiment, ten times reduction is used to expand the test image set to improve the expression recognition effect of the

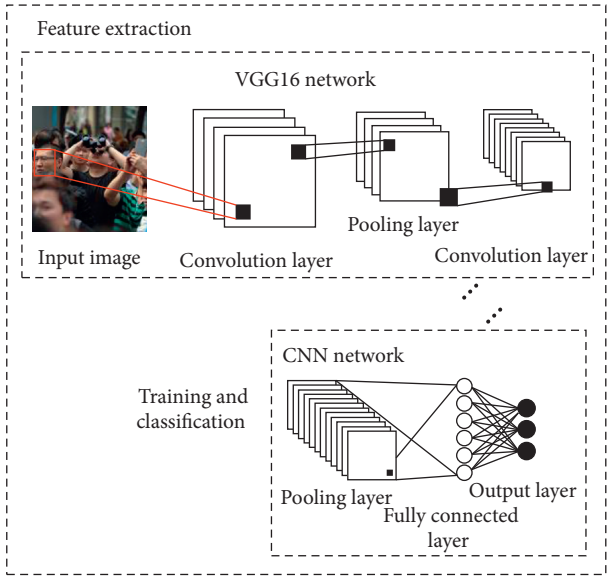

Figure 1: Overall framework of the proposed method.

network model, and experiments are used to verify the impact of ten times reduction on the expression recognition rate of the network model. Tenfold reduction refers to the reduction of the image in the upper left corner, lower left corner, upper right corner, lower right corner, and center, and the mirror operation is used at the same time. The tenfold reduction operation expands the number of input images by 10 times. The reduced image input model is tested, and the probability of 10 images output by the network model is calculated to obtain a mean value, which is used as the final classification result of the input image. This operation mode can effectively reduce the classification error rate of the network model.

The original input size of VGG16 network model is 224×224. If the image resolution is too small, VGG16 will not be able to fully extract feature information in feature extraction process, and if the image resolution is too large, it will increase the computational burden of computer memory during network model training [21, 22]. Therefore, in order to maintain the image resolution, the computer memory load is too heavy, resulting in the failure of the experiment. The proposed method enlarges the facial expression image of the dataset to 128×128, and then randomly reduce the image to 112×112. Finally, the image is randomly mirrored and input into the network model for training.

3.3. Improved VGG16 Model Feature Extraction. In the VGG16 network, due to the continuous use of small convolution cores for many times and the multiple growth of convolution cores in each layer, the number of corresponding output feature maps becomes more and occupies more graphics card space [23]. Experiments on the original model VGG16 show that a very large amount of parameters will be generated on the first fully connected layer, which makes the amount of calculation huge and consumes more computing resources. In addition, due to the size constraints of the dataset, the small- and medium-sized data samples do not perform well in the deep network, and the final experimental results are far lower than expected. Part of the reason is the overfitting problem caused by the small data scale, which leads to the insufficient generalization ability of the model and fails to reflect the original excellent performance of the depth network VGG16. Reducing the depth of the neural network in different ways to reduce the amount of parameters is helpful to prevent overfitting to a certain extent [24].

Inspired by GoogleNet and AlexNet, large convolution is used to reduce the dimension directly on the high-level feature map, which does not produce too much calculation. Moreover, the continuous large convolution kernel replaces the small convolution to reduce the complexity of the model, further compress the number of parameters, and reduce part of the full-connection layer, which will not affect the expression of the feature layer, but reduce the number of parameters [25]. Therefore, VGG16 structure is improved from two aspects: (1) The 5×5 convolution kernel is used on the larger feature map of the initial layer, and the 3×3 convolution kernel is used on the convolution layer stacked in the later three layers, which effectively reduces the space occupied by the feature map and maintains the feature extraction ability of the model; (2) Delete the fully connected layer 1 is deleted and directly connected with the fully connected layer 2. Secondly, the number of neurons in the remaining two fully connected layers is reduced to 1024 and 256. While reducing the number of parameters, it can make the features obtained by the last convolution layer more distinctive, which is helpful to improve the fusion effect. The improved VGG16 network constructed by aforementioned method is shown in Figure 2.

The input data dimension of the network is $224 \times 224 \times 3$. In the network, the rectified linear unit (ReLU) is used as the activation function, and dropout is used to solve the overfitting problem. Softmax is selected as the classifier for the classification task to estimate the probability of each class label in class K. Usually, the convolutional neural network needs to connect the low-dimensional full-connection layer after the convolution layer as a new feature layer to reduce the feature size, and the features obtained by the convolution layer usually contain rich image detail information. Therefore, the features obtained from the convolution are used as the features to be fused. Due to the large amount of VGG16 network parameters and easy overfitting, the proposed method uses large convolution kernel instead of small ones to reduce the complexity of the model, and reduces some fully connected layers to reduce the amount of parameters, so as to ensure the recognition efficiency on the basis of ensuring the accuracy.

3.4. Network Training. The whole network training takes the feature map extracted by the improved VGG16 network as a black box operation. The input layer data of CNN are the high-dimensional convolution abstract feature data output from the improved VGG16 network. The convolution neural

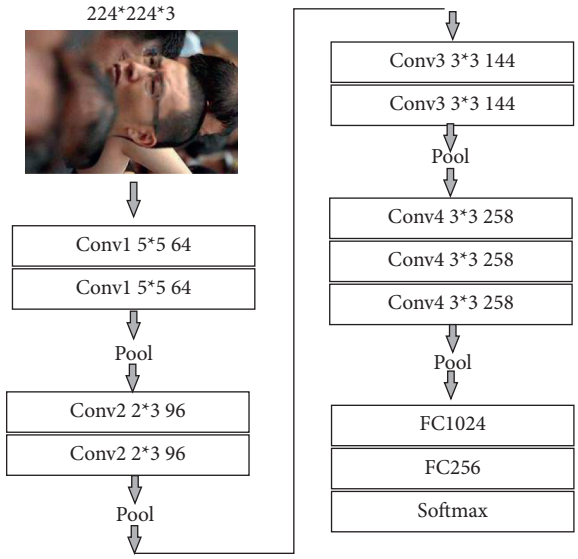

Figure 2: Improved VGG16 network.

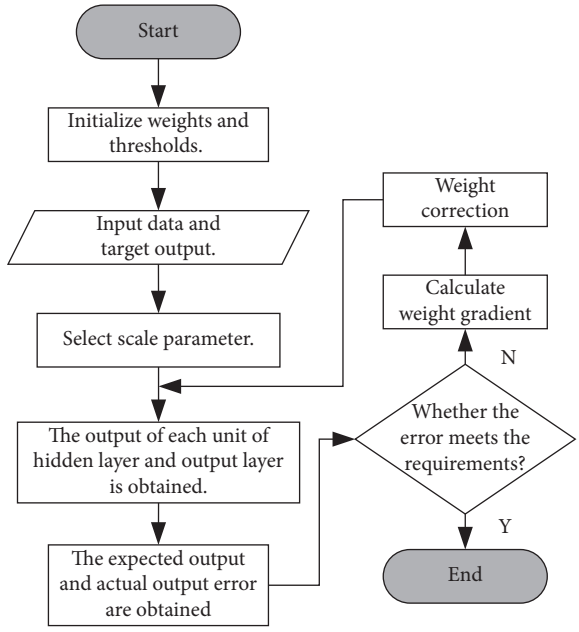

Figure 3: Overall training process of facial expression recognition networks.

network is used to train the characteristic data to obtain the training model [26]. The full-connection layer of CNN is a feedforward neural network. Its main responsibility is to undertake expression classification, but Softmax classifier is usually used to further improve the accuracy of expression classification. The overall training process of facial expression recognition network model is shown in Figure 3.

(1) Initialization and thresholds: There are seven categories of facial expressions. Therefore, compared with the traditional binary classification, the model parameters will increase significantly. The prediction model is a polynomial with parameters, and the prediction effect depends on the parameters of the model [27]. At present, the initialization of network parameters is mostly directly set to 0, which will lead to the same change trend of network parameters in the process of training, and even the parameter values are the same. Therefore, the proposed method uses random function assignment for the parameter initialization of the network model, which can avoid the influence of the initial parameter setting on the training model.

(2) Scale parameters: In machine learning, if the training data is less than the test data or the whole dataset itself is less, the whole model will have the phenomenon of overfitting or underfitting, resulting in poor prediction ability and failure to achieve the preset effect. In order to solve this imbalance, the corresponding proportion is selected by means of cross validation, retention method, and so on. In the training process, for the selection of network parameters, dropout is used to discard the parameters in proportion, so that the whole network becomes sparse, so as to enhance the generalization ability of the model. However, in the testing phase, dropout will be set to 0, that is, the parameters during training the model will be retained and used for model testing.

(3) Parameter iterative optimization is an important part of deep learning. When the gradient descent algorithm is used alone and when the initial value is far away from the minimum value, its step size will increase, and when it is close to the minimum value, its step size will become very slow [28]. When the gradient descent algorithm is used alone, regardless of the position where the initial value is far from or close to the minimum value, the step size will be large, and it is possible to skip directly for the next iteration. The proposed method combines the two methods and combines the gradient descent method and Gauss–Newton method through a linear combination. Its advantage is that it can quickly solve the problem of falling into local optimization and speed up the training speed.

3.5. Robot Interaction Design. The NAO robot uses NAOqi system. During program execution, a broker is used to load the required library files. Each library file contains one or more modules, such as "ALVideoDevice" vision module, "ALAudioDevice" sound module, and "ALMtion" action module.

3.5.1. Visual Module. The NAO robot has two 2D cameras, which are distributed on the forehead and mouth of the head. The camera can obtain pictures and video streams with a maximum resolution of 1280×960, which can be widely used in education, auxiliary medical treatment, and other fields. The camera adopts MT9M114 image sensor, and the photo format directly output by this device is YUV. In facial expression recognition, RGB color space is used. Therefore, it is necessary to subscribe to the "ALVideoDevice" module to change the YUV format into RGB. Considering the

limited CPU computing power of the NAO robot, the resolution of the proposed method is 640×480. This resolution will not cause too much loss of face image information, nor overload the processor of the robot.

3.5.2. Voice Module. The head of the NAO robot is equipped with four microphones and two speakers. The receiving frequency range of the microphone is 150 Hz–12 kHz, and the saved file format is WAV or OGG. In the NAOqi system, the "ALAudioDevice" sound module manages the input and output of audio. Therefore, when you need to read the data of the microphone, you must subscribe to the "ALAudioDevice" sound module.

3.5.3. Action Module. In the process of robot interaction, the face images collected by the NAO camera are output to the NAO robot after model prediction at the computer ends. The robot makes different interactive actions according to different expressions. The "ALMtion" action module includes methods related to the robot's actions. The robot action is written with the official Choregraphe graphical programming software. Using the time axis command box to design the formulated action, it is not necessary to use complex robot kinematics knowledge and program programming. The time axis is shown in Figure 4.

4. Experiment and Analysis

In the experiment, the hardware environment is one desktop computer, the CPU is Intel (R) core (TM) i7, 16 GB memory, and the GPU is GTX 1060. The development environment is TensorFlow framework based on Python language. In the training process, all learning rates are set to 0.01, the batch size is 500, the momentum is 0.99, the weight attenuation is 0.0001, the number of iterations is set to 100, and the maximum limit number is 10000. At the same time, the random gradient descent optimization method is used to train the network.

In addition, the data set used in the experiment is CK + data set. CK + dataset is an extension of Cohn-Kanade dataset, which is commonly used in facial expression recognition research. This dataset was released in 2010 and can be obtained free of charge. It is often used in facial expression recognition research. This dataset collected the frontal facial expressions of 123 people, a total of 593 image sequences. There are 7 facial expressions in total, including anger, disgust, fear, happiness, sadness, surprise, and contempt. The picture example of CK + dataset is shown in Figure 5.

988 pictures were selected, including 136 angry, 178 disgusted, 76 afraid, 208 happy, 85 sad, 250 surprised, and 55 neutral. 14 anger, 10 disgust, 11 fear, 23 happiness, 11 sadness, 26 surprise, and 8 neutral vision were selected as the test set, and the rest as the training set.

4.1. Training and Testing Accuracy versus Cycle Curve. The facial expression recognition accuracy curve of VGG16 network model and the proposed improved VGG16 network model during training and testing on the facial expression dataset CK + is shown in Figure 6.

It can be seen from Figure 6 that the improved VGG16 achieves 100% training accuracy faster during training, which proves that the fitting speed of the proposed network is faster. The improved VGG16 network has higher expression recognition accuracy than VGG16 network. Taking the test set as an example, the recognition accuracy of the improved VGG16 network is about 90%, which proves that its fitting effect is also better. In order to further improve the recognition accuracy, the proposed method combines the feature extraction advantages of improved VGG16 network and the classification characteristics of convolutional neural network to realize the high-precision human-computer interaction of the NAO robot system.

4.2. Influence of Different Feature Dimensions on Recognition Effect. In order to verify the influence of different networks and different feature dimensions on the recognition effect, the VGG16 network and the improved VGG16 network are used to conduct experiments on CK + dataset, and the features of different dimensions are selected to compare the recognition accuracy. The results are shown in Table 1.

As can be seen from Table 1, with the increase of feature dimension, the recognition accuracy of the two networks does not increase, but decreases slightly at 256 dimensions, indicating that the feature dimension is sufficient to characterize the CK + dataset at 256 dimensions, which contains most of the effective information in the dataset. The texture and detail features extracted by the proposed improved VGG16 in the shallow network are richer than those in the VGG16, especially some key features, such as eye feature information. In the deeper network, the improved VGG16 extracts more features such as contour and shape, especially the abstract features obtained in the last convolution layer, which are relatively more representative. At the same time, the improved VGG16 network does not reduce the ability of feature extraction due to the simplified structure. Therefore, this paper adopts the 256 dimensional feature dimension.

4.3. The Confusion Matrix. In the facial expression dataset CK + , the confusion matrix of 8 kinds of expression recognition results by the VGG16 network before and after improvement is shown in Figure 7.

As can be seen from Figure 7, compared with the VGG16 network, the improved VGG16 network improves the facial expression recognition rate of facial expression. In the original model, the highest accuracy is the recognition accuracy of happy expression, which reaches 92.83%. The lowest accuracy rate was neutral expression, only 71.16%; the average recognition accuracy is 84.03%. The improved VGG16 network optimizes the fully connected layer, so that the features obtained by the last layer of convolution are more differentiated, so the recognition effect is better. In addition, for the seven kinds of facial expressions, the recognition accuracy of disgust, neutral, and other expressions is low and easy to be confused. However, the improved VGG16 network uses a continuous large convolution kernel

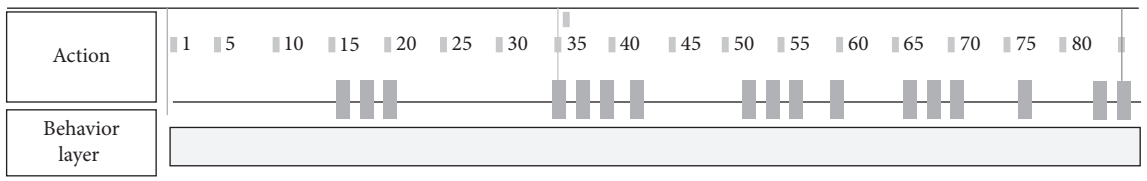

Figure 4: Time axis.

Figure 5: Picture example of CK + dataset.

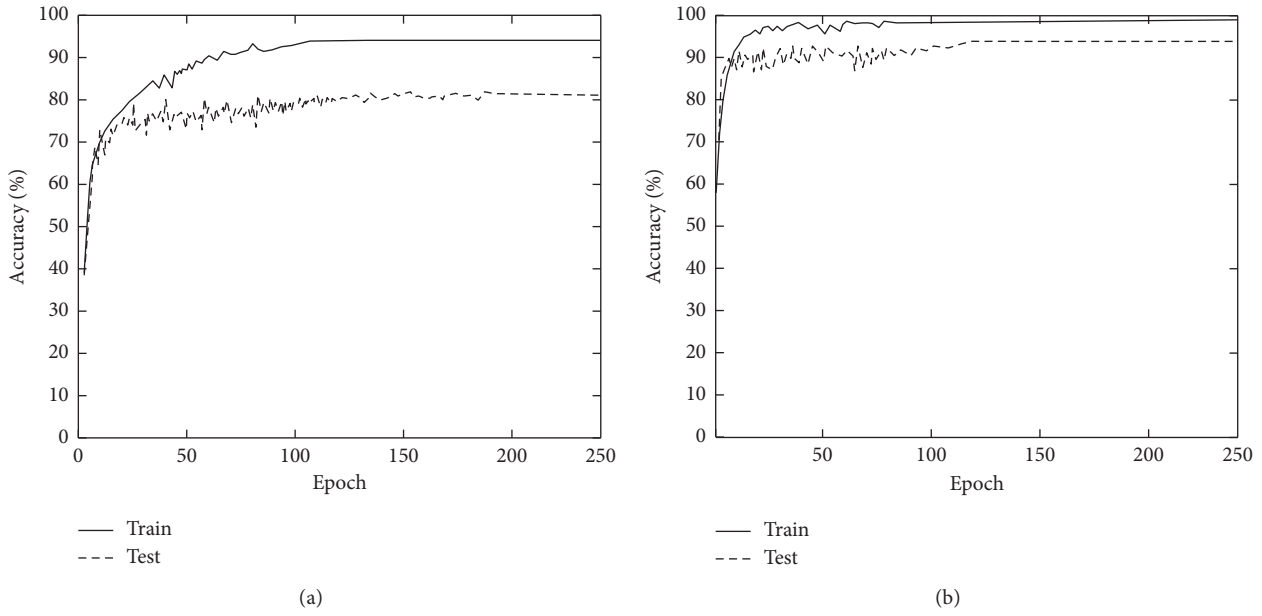

Figure 6: Expression recognition accuracy during training and testing. (a) VGG16 network training accuracy. (b) Improved VGG16 network training accuracy.

instead of a small convolution kernel and reduces part of the full-connection layer, which can better extract the expression detail features. Therefore, its recognition accuracy is better than the VGG16 network. The highest accuracy of the improved model is the recognition of happy expression, which is 98.78%; the lowest was disgusting expression, and

Table 1: Recognition accuracy of different feature dimensions.

Network structure	Feature dimension (%)			
	64	256	512	1024
VGG16	86.73	85.94	81.59	73.28
Improved VGG16	90.51	89.27	85.12	80.35

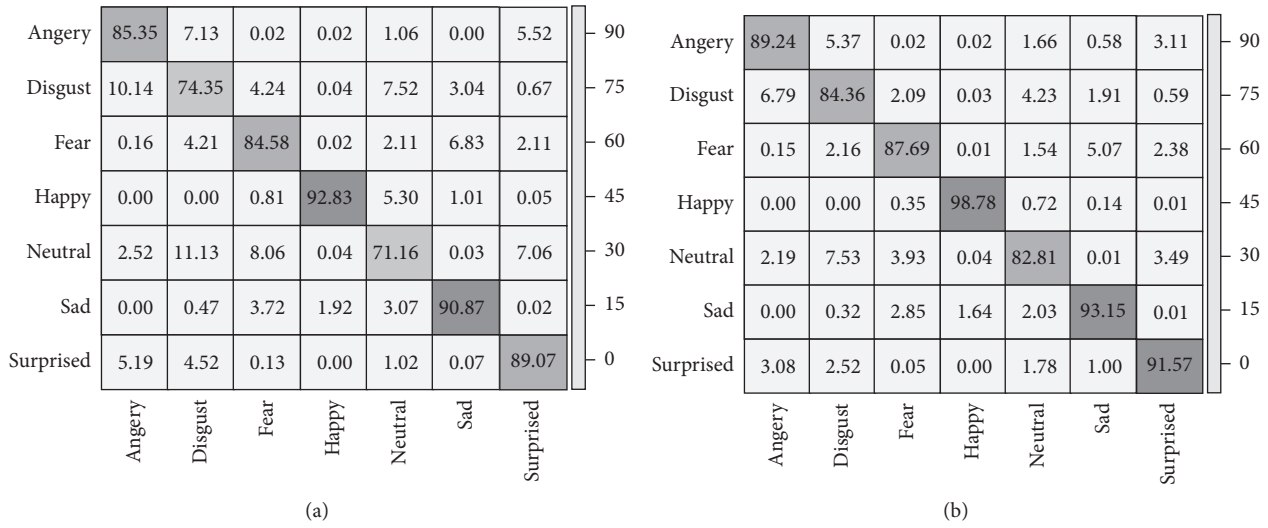

Figure 7: Confusion matrix of expression recognition results of different networks. (a) The VGG16 network. (b) The improved VGG16 network.

the recognition accuracy was 84.36. The average recognition accuracy is 89.66%, which is 5.63% higher than that of the original model. Therefore, it can be explained that the performance of the improved VGG16 network model is better than that of the traditional VGG16 network.

4.4. Human-Computer Interaction Test. The trained facial expression model is completed in Python 3.6, while the NAO robot can only support Python 2.7, so it cannot be deployed directly. Therefore, this interactive experiment is conducted through wireless connection. During the connection process, it should be noted that when the robot connects with the computer for the first time, it needs to insert the network cable to obtain the corresponding IP address, then log in to the web page through the IP address and set up the router for wireless connection. After the robot is connected to the network normally, the program is downloaded to the robot wirelessly. The PC obtains the face image by calling the robot's camera and sends it to the trained model. The predicted results of the model are then sent back to the NAO robot. The robot makes corresponding voice and actions according to the recognized emotions. This paper selects two representative expressions: happy and sad. The interaction results are shown in Figure 8.

As can be seen from Figure 8, the picture on the left is the expression result predicted by the model after the camera acquisition. The picture on the right shows the voice and corresponding actions made by the robot. In order to visualize the voice effect, the virtual robot was connected and the voice results were directly visualized. When the happy

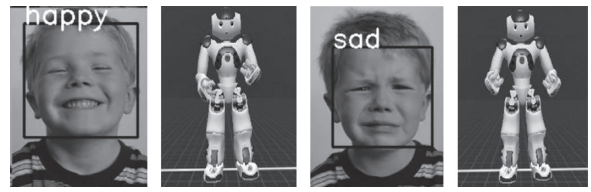

Figure 8: The emotional interaction results.

expression is detected, the robot will interact with you and say "You look happy. Can you share it with me?" When the sad expression is detected, the robot will make a stand up action and say "I'll be unhappy if you're sad." Through many tests, the interaction results show that the robot can stably recognize emotions and can make corresponding action interaction.

5. Conclusions

Face recognition technology has been relatively mature and has been widely used in many fields. However, face expression recognition is still in an exploratory stage. Aiming at the problems of insufficient feature data and a low recognition rate in face expression recognition, an expression recognition method based on the improved VGG16 network model is proposed. The improved VGG16 network is used to extract the features of facial expression gray image, and the high-level features are input into the CNN model. The classification of facial expression is realized by Softmax classifier. The comprehensive experimental results show that

the improved VGG16 network model is effective in training and testing expression recognition. With the expansion of the depth and structure of the network, the application of certain networks in some fields has shown better results. The next step of in-depth research can also consider the use of deeper network structures, structural innovations, and the parallel connection of multiple networks. And other methods are applied to the facial expression recognition system.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

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

Acknowledgments

This work was supported by the scientific research project from Hunan Provincial Department of Education (no. 20C0218).

References

[1] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognition Letters*, vol. 120, no. 8, pp. 69–74, 2019.

[2] Y. Chen, Z. Zhang, L. Zhong, T. Chen, J. Chen, and Y. Yu, "Three-stream convolutional neural network with squeeze-and-excitation block for near-infrared facial expression recognition," *Electronics*, vol. 8, no. 4, pp. 385–397, 2019.

[3] Y. Gan, J. Chen, and L. Xu, "Facial expression recognition boosted by soft label with a diverse ensemble," *Pattern Recognition Letters*, vol. 125, no. 7, pp. 105–112, 2019.

[4] X. Fan and T. Tjahjadi, "Fusing dynamic deep learned features and handcrafted features for facial expression recognition," *Journal of Visual Communication & Image Representation*, vol. 65, no. 11, pp. 102659.1–102659.6, 2019.

[5] D. Li, X. Zhao, G. Yuan, and X. Zhao, "Robustness comparison between the capsule network and the convolutional network for facial expression recognition," *Applied Intelligence*, vol. 51, no. 4, pp. 1–10, 2021.

[6] S. Cao, Y. Yao, and G. An, "E2-capsule neural networks for facial expression recognition using AU-aware attention," *IET Image Processing*, vol. 14, no. 11, pp. 2417–2424, 2020.

[7] A. Renda, M. Barsacchi, A. Bechini, and F. Marcelloni, "Comparing ensemble strategies for deep learning: an application to facial expression recognition," *Expert Systems with Applications*, vol. 136, no. 11, pp. 1–11, 2019.

[8] F. Wang, J. Lv, G. Ying, S. Chen, and C. Zhang, "Facial expression recognition from image based on hybrid features understanding," *Journal of Visual Communication and Image Representation*, vol. 59, no. 3, pp. 84–88, 2019.

[9] W. Cao, Z. Feng, D. Zhang, and Y. Huang, "Facial expression recognition via a CBAM embedded network," *Procedia Computer Science*, vol. 174, no. 4, pp. 463–477, 2020.

[10] N. Amjed, F. Khalid, R. W. O. K. Rahmat, and H. B. Madzin, "A robust geometric skin colour face detection method under unconstrained environment of smartphone database," *Applied Mechanics and Materials*, vol. 892, no. 3, pp. 31–37, 2019.

[11] H. Ma and T. Celik, "FER-Net: facial expression recognition using densely connected convolutional network," *Electronics Letters*, vol. 55, no. 4, pp. 184–186, 2019.

[12] M. Sajjad, S. Zahir, A. Ullah, Z. Akhtar, and K. Muhammad, "Human behavior understanding in big multimedia data using CNN based facial expression recognition," *Mobile Networks and Applications*, vol. 25, no. 4, pp. 1611–1621, 2020.

[13] K. Li, Y. Jin, M. W. Akram, R. Han, and J. Chen, "Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy," *The Visual Computer*, vol. 36, no. 2, pp. 391–404, 2020.

[14] C. Xu, Y. Cui, Y. Zhang, P. Gao, and J. Xu, "Person-independent facial expression recognition method based on improved Wasserstein generative adversarial networks in combination with identity aware," *Multimedia Systems*, vol. 26, no. 1, pp. 53–61, 2020.

[15] H. Zhang, Z. Qu, L. Yuan, and G. Li, "A Face Recognition Method Based on LBP Feature for CNN[C]," in *Proceedings of the 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference*, pp. 544–547, IEEE, Chongqing, China, March 2017.

[16] G. Yolcu, I. Oztel, S. Kazan et al., "Facial expression recognition for monitoring neurological disorders based on convolutional neural network," *Multimedia Tools and Applications*, vol. 78, no. 22, pp. 31581–31603, 2019.

[17] J. Shao and Y. Qian, "Three convolutional neural network models for facial expression recognition in the wild," *Neurocomputing*, vol. 355, no. 8, pp. 82–92, 2019.

[18] H. Wang, J. Lu, and L. Nwosu, "Two-channel convolutional neural network for facial expression recognition using facial parts," *International Journal of Big Data Intelligence*, vol. 6, no. 3/4, pp. 259–268, 2019.

[19] M. Aamir, T. Ali, A. Shaf, M. Irfan, and M. Q. Saleem, "ML-DCNNet: multi-level deep convolutional neural network for facial expression recognition and intensity estimation," *Arabian Journal for Science and Engineering*, vol. 45, no. 12, pp. 10605–10620, 2020.

[20] Y. Ye, X. Zhang, Y. Lin, and H. Wang, "Facial expression recognition via region-based convolutional fusion network," *Journal of Visual Communication and Image Representation*, vol. 62, no. 7, pp. 1–11, 2019.

[21] J. Li, Y. Mi, and G. Li, "CNN-based facial expression recognition from annotated RGB-D images for human–robot interaction," *International Journal of Humanoid Robotics*, vol. 16, no. 4, pp. 504-505, 2019.

[22] D. K. Jain, Z. Zhang, and K. Huang, "Multi angle optimal pattern-based deep learning for automatic facial expression recognition," *Pattern Recognition Letters*, vol. 139, no. 3, pp. 157–165, 2020.

[23] U. B. Chavan and D. Kulkarni, "Optimizing deep convolutional neural network for facial expression recognition," *European Journal of Engineering Research and Science*, vol. 5, no. 2, pp. 192–195, 2020.

[24] R. Ramya, K. Mala, and S. Selva Nidhyananthan, "3D facial expression recognition using multi-channel deep learning framework," *Circuits, Systems, and Signal Processing*, vol. 39, no. 2, pp. 789–804, 2020.

[25] W. Wei, Q. Jia, Y. Feng, G. Chen, and M. Chu, "Multi-modal facial expression feature based on deep-neural networks," *Journal on Multimodal User Interfaces*, vol. 14, no. 1, pp. 17–23, 2020.

[26] Y. Cai, J. Gao, G. Zhang, and Y. Liu, "Efficient facial expression recognition based on convolutional neural network," *Intelligent Data Analysis*, vol. 25, no. 1, pp. 139–154, 2021.

[27] S. Suchitra, P. S. Sathya, P. Balachandran, and M. Faustina, "Intelligent driver warning system using deep learning-based facial expression recognition," *Scopus*, vol. 8, no. 3, pp. 831–838, 2019.

[28] A. Caroppo, A. Leone, and P. Siciliano, "Comparison between deep learning models and traditional machine learning approaches for facial expression recognition in ageing adults," *Journal of Computer Science and Technology*, vol. 35, no. 5, pp. 1127–1146, 2020.