Scholar Publishing Group
International Journal of Neural Network
https://doi.org/10.38007/NN.2020.010202
ISSN 2790-2692 Vol. 1, Issue 2: 9-16

Facial Expression Recognition Based on Neural Network and Feature Extraction

Keyi Yin*

Zhengzhou Normal University, Zhengzhou, China

**corresponding author*

Keywords: Neural Network, Feature Extraction, Facial Expression, Expression Recognition

Abstract: Facial expression recognition has always been an interesting and challenging problem, which is easily affected by various factors in practical applications, such as illumination, posture, facial occlusion, age, race and so on. This paper focuses on facial expression recognition based on neural network and feature extraction. In this paper, a new convolutional neural network model for facial expression recognition is proposed to solve the problem that the traditional convolutional neural network is too complex. In the initial stage, the interleaved group convolution structure is adopted, which effectively reduces the number of channel connections between network layers, and at the same time, the interleaved structure also enables effective information communication between different volume units. Through verification and comparison on CK+ and FER datasets in this paper, the improved convolutional neural network facial expression recognition model proposed in this paper can better complete the task of facial expression recognition.

1. Introduction

With the substantial improvement of computing capability of computer hardware, artificial intelligence and deep learning technology have been applied to all aspects of life. Mini robots that automatically deliver food in restaurants; There are intelligent voice assistants in mobile phones and computers, which can complete the functions of making calls and sending messages without hands; In terms of entertainment, virtual reality wearable devices are increasingly applied in games, bringing people new game experience [1-2]. With the continuous improvement of people's quality of life, there will be higher standards in the use of human-machine interaction experience, and robots that can interact with people more intelligently will have greater demand. Therefore, how to understand human emotions and interact with people naturally is the inevitable direction of future robot development. Facial expression activity is an important part of People's Daily communication.

In daily life or work, when people communicate, they will be accompanied by some expressions. To convey non-verbal information through facial expressions, thereby helping others to understand themselves[3-4]. Facial expression is a very important way of information transmission, if the computer can accurately recognize and analyze the facial expression, it can extract a lot of valuable information. As a key technology of emotion computing system, expression recognition has a broad application prospect. With the development of artificial intelligence, the traditional keyboard mouse or touch screen interaction can no longer meet people's needs. Under the new technological background, people's life needs more convenient human-computer interaction [5]. Facial expression recognition is a very important part of human-computer interaction and a problem that must be solved to realize a new interaction mode. By sensing the emotional tendency of the service object and further analyzing its demand intention, the computer can realize a more intelligent and humanized human-computer interaction experience [6].

With the improvement of hardware computing power and the support of big data, deep learning technology has been widely applied and developed in various fields. Deep learning has achieved significant results in face recognition, target detection, action estimation and other aspects, which has certain reference significance for facial expression recognition [7]. Convolutional Neural Network (CNN) has a good effect in image classification and recognition. At the initial stage, researchers only used CNN as a feature extractor, instead of manually extracted features, and still used traditional machine learning algorithms for classification and recognition. Using CNN for facial expression recognition, most of the algorithms proposed in current studies are "end-to-end" processing, which complete feature extraction, classification and recognition in a network [8]. Although there is a deep learning technique is applied to facial expression recognition has achieved good results, but as the problem is more and more close to actual application scenario, the challenge for the more and more complex, not only need reliable recognition accuracy, in the actual deployment model, response time, running memory consumption and other issues need to be taken into consideration when it is applied in actual [9]. Therefore, the problem of facial expression recognition is still a very challenging topic to be explored by researchers.

Combined with a new convolution operation, this paper proposes a novel convolutional neural network model, which reduces the number of pooling layers in the model and ensures the receptive field range of feature map. This model is more suitable for facial expression recognition of low-resolution images, and can effectively improve the effect of facial expression recognition.

2. Facial Expression Recognition Based on Convolutional Neural Network

2.1. Image Feature Extraction

With the amazing and rapid development of convolutional network methods based on deep learning in all directions of computer vision tasks, the accuracy of manual classification has been surpassed by deep learning methods, even for solving some problems such as image classification of large-scale samples. It has been demonstrated that the convolutional neural network is very powerful in image classification and processing. However, without the help of current high-performance computers, previous researchers' understanding of CNN in deep learning methods was only a relatively complex mapping relationship in mathematical problems [10-11]. Although CNN can input an image into the network model, and then carry out superposition mapping mode through the input layer, hidden layer and output layer to form an abstract space for feature classification. However, they do not know the actual specific process of feature extraction by CNN method in images, and in the process of increasing depth of deep convolutional network, the scale of CNN model becomes larger and larger, which makes the training and learning process very difficult [12-13]. In order to construct and realize efficient convolution model, more scholars began to study the principle of convolution model in the operation and processing process. With a reverse convolution model proposed by various scholars, it can turn each layer of feature information obtained in the process of CNN convolution into the output in the form of images, and we begin to understand the operation and processing process of CNN more clearly [14].

In the process of image feature extraction by CNN, the convolution kernel in each convolutional layer is equivalent to a feature extractor, and the feature information extracted by the convolutional kernel at different levels is different [15]. Meanwhile, the convolutional kernel in the first few layers of CNN extracts the lowest image features such as color features and texture features. However, when the number of convolutional layers increases continuously, the range of feature images becomes smaller and smaller, and the receptive field obtained by the convolutional layer begins to grow larger and larger [16]. Through the continuous superposition of convolutional layers, the convolutional kernel gradually begins to extract more complex and more recognizable abstract features.

The image features extracted by the modules such as convolution, pooling and full connection in the CNN model are all first-order. In the rough image classification task, due to the great differences in image features of different basic categories, good classification results can be obtained by using the first-order features of images [17]. However, when solving the problem of image expression classification, there are only slight differences between different types of expressions, and the accuracy rate is not that high when we classify expressions only by first-order features [18].

2.2. Construct an Improved Convolutional Neural Network Model

The network proposed in this paper follows the consistent idea of lightweight network "less channels with more layers", and the network is divided into three stages. In the first stage, the interleaved group convolution structure is used to reduce the number of channels connected in the network layer and communicate with each other to increase the information fault tolerance rate of the network. In the second stage, the separable convolution structure is combined with the residual network to optimize the feature extraction performance while reducing the parameters. In the third stage, the Softmax function is used to obtain the final classification results.

Stage 1: The front-end position of the CNN will simultaneously acquire positive and negative phase information in the network structure, while the cascaded linear rectifier unit (ReLU) will delete the front-end information, resulting in a large amount of redundant information generated by the convolution kernel. The parameters distributed at the front of the network have a strong negative correlation, which gradually weakens with the deepening of the network layers. Therefore, we adopt CReLU as the activation function of the front-end network. The expression for CReLU is as follows:

$$C\mathrm{Re}LU(x) = [\mathrm{Re}LU(x), \mathrm{Re}LU(-x)] \tag{1}$$

According to the expression, the output dimension of CReLU is automatically doubled. According to this principle, CReLU reduces the output channels by half and cascades the remaining channels with the negative output in a simple concatenation manner. Therefore, CReLU enables the front-end network to achieve a 2x speed increase without loss of accuracy.

The overall framework of the first stage adopts interleaved group convolution structure, with group convolution as the basic structure of this part. This structure divides the input layer and output layer into corresponding partitions, which have no relationship with each other, and the input layer is only convolved with the corresponding partitions of the output layer. This operation will effectively reduce the number of channels between the connecting layers. Assuming that the input layer and the output layer are divided into G intervals, the theoretical computation of this part of the network will be reduced to 1/G of the original. Although the group convolution can effectively reduce the computation amount, the division of the interval will also cause the loss of information communication between channels, so the interleaved structure is added on the basis of the group convolution.

Second stage: Inspired by the proposed lightweight network, the network model is a structure composed entirely of convolution, and the output of interleaved group convolution is used as the input of the second stage, which combines the residual structure with the depth-separable convolution. By using point-by-point convolution, the number of feature channels is firstly increased and then decreased, so that the deep convolutional layer works in high-dimensional features, and the negative response input and output from ReLU to the front end are close to zero. Moreover, some functions are lost due to compression when ReLU is transferred. It adopted Linear Linear ReLU bottlenecks to replace the original.

Stage 3: After four depth-separable residual convolution modules, a conventional convolution layer is added, followed by a random connection layer and a global average pooling layer to reduce the number of features and regularize them to prevent over-fitting. Finally, they are output in the form of vectors. The dimension represents the number of expressions. Therefore, a Softmax layer is added at the bottom for the final expression classification.

$$Soft\max = S_j = \frac{1}{\sum_{j=1}^{T} e_j x^{(i)}} \begin{bmatrix} e^{\theta_1 x^{(i)}} \\ e^{\theta_2 x^{(i)}} \\ \cdots \\ e^{\theta_T x^{(i)}} \end{bmatrix} \tag{2}$$

Where, Sj is the normalized probability distribution, θ represents the parameters of the model, x is the input features of the classification layer, and T is the total number of classifications.

Its loss function is as follows:

$$Loss = -\sum_{j=1}^{T} y_i (S_j) \tag{3}$$

yi is a vector of 1×T, among T values, one and only one value is 1(the position corresponding to the true label), and the others are all 0. Part of the structure retains the initial parameters of the original network.

3. Simulation Experiment of Convolutional Neural Network Model

3.1. TensorFlow Platform

All experiments in this paper are based on Tensorflow deep learning platform. TensorFlow is an open source software library that uses data flow graph for numerical calculation. It is the second generation of artificial intelligence research and development system developed by Google. It transfers complex data structures to artificial neural networks for analysis and processing, and supports convolutional neural networks, recurrent neural networks and other deep neural network models. The system can be used in many deep learning areas, and has been widely used in Google products and services. It has been deployed in more than 100 machine learning projects, covering more than a dozen areas such as speech recognition, computer vision, robotics, information retrieval, information extraction, natural language processing, and drug testing.

The experimental environment is 64-bit Microsoft Windows 10 operating system, GPU is NVIDIA GeForceMX250, and the memory is 16GB. TensorFlow is 1.4.0, CUDA is 8.0.44, CUDNN is 6.0, and Keras is 2.1.

3.2. Experimental Data Set

In this paper, 2863 CK+ data sets and 34515 FER data sets were collected. In order to test the effectiveness of the algorithm, the experiment adopts the 50% fold cross validation method. The dataset was randomly divided into five parts, four of which were used as training and the other one was used as testing. Five experiments were conducted, and the final result was the average of the five experimental results.

4. Simulation Experiment Results

4.1. CK+ Data Set

Table 1. Comparison of the accuracy of four expressions in CK+ dataset

	Happiness	Sadness	Fear	Anger
CNN	0.99	0.99	1.00	0.93
Improve the CNN	1.00	0.99	1.00	0.98

Table 2. Comparison of recall rate of four emoticons in CK+ dataset

	Happiness	Sadness	Fear	Anger
CNN	1.00	1.00	0.97	0.96
Improve the CNN	1.00	1.00	0.99	0.97

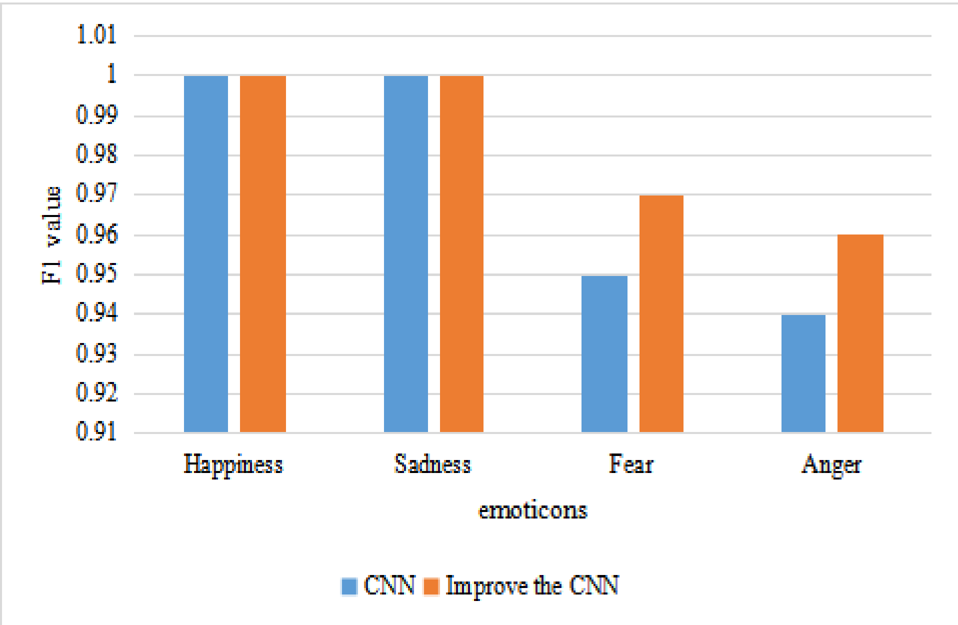

Figure 1. Comparison of F1 values of four expressions in CK+ dataset

As shown in Table 1, Table 2 and Figure 1. Is the performance comparison between CNN and the improved CNN model proposed in this paper on CK+ dataset. In the CK+ dataset, the accuracy rate, recall rate and F1 score of expressions "happy", "sad" and "afraid" reach or are infinitely close to 1, indicating that these three expression datasets have obvious features, and the three integrated models have high applicability and accurate prediction. The accuracy rate of expressions "disgust" and "fear" is 1, and the recall rate is not ideal, because the image noise problem exists in the sample, and there is unnecessary or redundant interference information in the image data.

4.2. FER Data Set

Table 3. Comparison of expression accuracy in FER dataset

	Happiness	Sadness	Fear	Anger
CNN	0.83	0.65	0.54	0.57
Improve the CNN	0.86	0.66	0.63	0.65

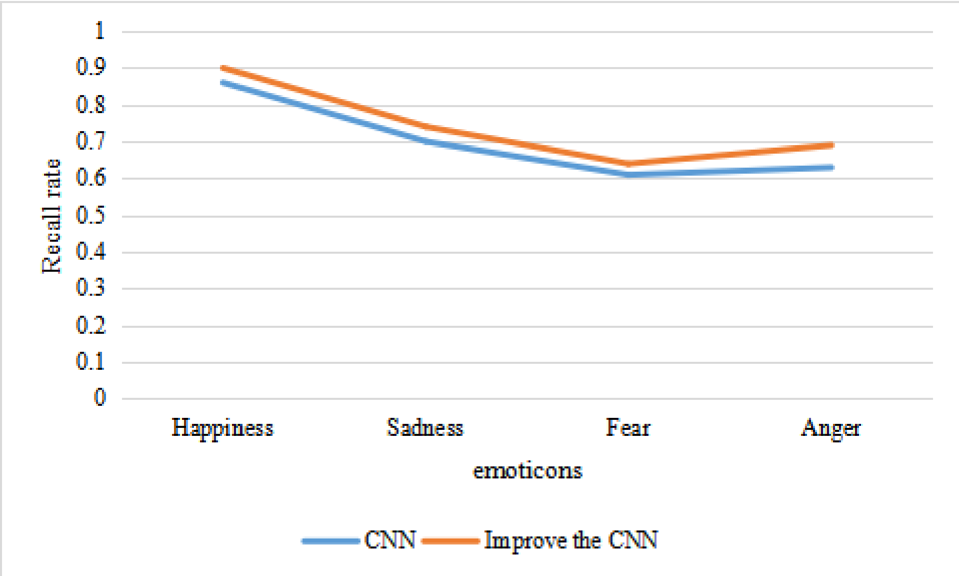

Figure 2. Comparison of recall rate of seven emoticons in FER dataset

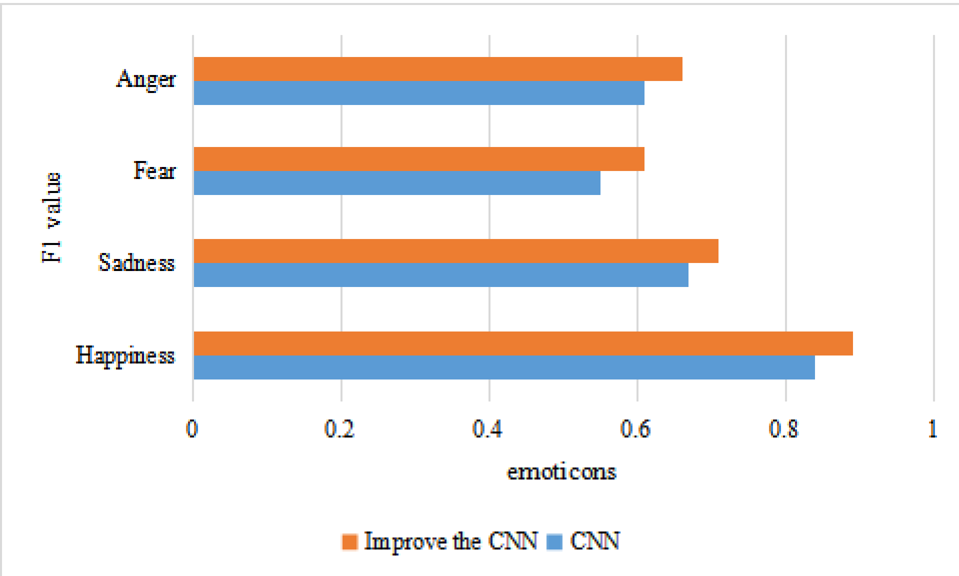

Figure 3. Comparison of F1 value of seven expressions in FER dataset

See Table 3, Figure 2 and Figure 3. Is the performance comparison of CNN and the improved CNN model proposed in this paper on FER dataset? In the FER dataset, the expression "happy" has the highest evaluation index, and this expression is also the most easily recognized by humans.

5. Conclusion

Facial expression contains abundant information, which is an important way to express emotion and communicate information. With the development of science and technology, more and more people try to use computer vision technology and image processing technology to realize the automatic recognition of facial expressions. Facial expression recognition has important research significance, which is embodied in the field of medicine, security, game and entertainment. However, facial expression recognition is a complex and challenging research direction. This paper first introduces the relevant theories of facial expression recognition and deep learning knowledge, secondly summarizes the facial expression recognition of specific steps, and then further study the convolutional neural network, the internal structure of the convolutional neural network structure is improved and based on the research of the integrated approach, after many experiments to find the optimal integration model, and applies this idea to the real-time identification system, The robustness of the system is improved effectively.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Nguyen H D, Yeom S, Lee G S, et al. Facial Emotion Recognition Using an Ensemble of Multi-Level Convolutional Neural Networks. International Journal of Pattern Recognition and Artificial Intelligence, 2018, 33(11):139-149. https://doi.org/10.1142/S0218001419400159

[2] Nishime T, Endo S, Toma N, et al. Feature Acquisition and Analysis for Facial Expression Recognition Using Convolutional Neural Networks. Transactions of the Japanese Society for Artificial Intelligence, 2017, 32(5): F-H34_1-8. https://doi.org/10.1527/tjsai.F-H34

[3] [1]Ruicong, ZHI, Hairui, et al. Combining 3D Convolutional Neural Networks with Transfer Learning by Supervised Pre-Training for Facial Micro-Expression Recognition. IEICE Transactions on Information and Systems, 2019, E102.D (5):1054-1064. https://doi.org/10.1587/transinf.2018EDP7153

[4] Lee J, Kim S, Kim S, et al. Multi-Modal Recurrent Attention Networks for Facial Expression Recognition. IEEE Transactions on Image Processing, 2020, PP(99):1-1. https://doi.org/10.1109/TIP.2020.2996086

[5] Pons G, Masip D. Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis. Affective Computing IEEE Transactions on, 2018, 9(3):343-350.

https://doi.org/10.1109/TAFFC.2017.2753235

[6] Suchitra S, Sathya P S, Balachandran P, et al. Intelligent Driver Warning System using Deep Learning-based Facial Expression Recognition. Scopus, 2019, 8(3):831-838. https://doi.org/10.35940/ijrte.C4028.098319

[7] Mohan K, Seal A, Krejcar O, et al. Facial Expression Recognition Using Local Gravitational Force Descriptor-Based Deep Convolution Neural Networks. IEEE Transactions on Instrumentation and Measurement, 2020, PP(99):1-1. https://doi.org/10.1109/TIM.2020.3031835

[8] R, Santhoshkumar, M, et al. Deep Learning Approach for Emotion Recognition from Human Body Movements with Feedforward Deep Convolution Neural Networks - ScienceDirect. Procedia Computer Science, 2019, 152(C):158-165. https://doi.org/10.1016/j.procs.2019.05.038

[9] Alam M, Vidyaratne L S, Iftekharuddin K M. Sparse Simultaneous Recurrent Deep Learning for Robust Facial Expression Recognition. IEEE Transactions on Neural Networks & Learning Systems, 2018, PP(99):1-12. https://doi.org/10.1016/j.neunet.2018.04.020

[10] Jain N, Kumar S, Kumar A , et al. Hybrid deep neural networks for face emotion recognition. Pattern Recognition Letters, 2018, 115(NOV.1):101-106. https://doi.org/10.1016/j.patrec.2018.04.010

[11] Raheem F A, Al A. Deep Learning Convolution Neural Networks Analysis and Comparative Study for Static Alphabet ASL Hand Gesture Recognition. Xi'an Dianzi Keji Daxue Xuebao/Journal of Xidian University, 2020, 14(4):1871-1881. https://doi.org/10.37896/jxu14.4/212

[12] Renda A, Barsacchi M, Bechini A, et al. Comparing ensemble strategies for deep learning: An application to facial expression recognition. Expert Systems with Application, 2019, 136(Dec.):1-11. https://doi.org/10.1016/j.eswa.2019.06.025

[13] Gupta O, Raviv D, Raskar R. Multi-Velocity Neural Networks for Facial Expression Recognition in Videos. Affective Computing, IEEE Transactions on, 2019, 10(2):290-296. https://doi.org/10.1109/TAFFC.2017.2713355

[14] Rodriguez P, Cucurull G, Gonalez J, et al. Deep Pain: Exploiting Long Short-Term Memory Networks for Facial Expression Classification. IEEE Transactions on Cybernetics, 2017, PP(99):1-11.

[15] Inthiyaz S, Parvez M M, Kumar M S, et al. Facial Expression Recognition Using KERAS. Journal of Physics Conference Series, 2020, 1804(1):012202. https://doi.org/10.1088/1742-6596/1804/1/012202

[16] Mary A H, Kadhim Z B, Sharqi Z S. Face Recognition and Emotion Recognition from Facial Expression Using Deep Learning Neural Network. IOP Conference Series: Materials Science and Engineering, 2020, 928(3):032061 (16pp). https://doi.org/10.1088/1757-899X/928/3/032061

[17] Lee H J, Lee D. Study of Process-Focused Assessment Using an Algorithm for Facial Expression Recognition Based on a Deep Neural Network Model. Electronics, 2020, 10(1):54. https://doi.org/10.3390/electronics10010054

[18] Jayasimha Y, Reddy R. A robust face emotion recognition approach through optimized SIFT features and adaptive deep belief neural network. Intelligent decision technologies, 2019, 13(3):379-390. https://doi.org/10.3233/IDT-190022