Adaptive Convolutional Neural Network-based Information Fusion for Facial Expression Recognition

Ting-Ting Gao, Hang Li, and Shou-Lin Yin
(Corresponding author: Hang Li)

Software College & Shenyang Normal University
253 Huanghe Bei Dajie, Shenyang City, Liaoning Province, China
(Email: 841584960@qq.com)

(Received Sept. 22, 2020; Revised and Accepted Dec. 4, 2020; First Online Dec. 12, 2020)

Abstract

Aiming at the problems of insufficient feature extraction and low recognition rate of convolutional neural network, an adaptive convolutional neural network-based information fusion for Facial expression recognition is proposed. In this paper, the method of gradient feature and texture feature fusion is adopted to effectively combine the local shape feature and local texture feature to perform a more comprehensive representation of facial expression information and obtain different feature information for information fusion at the full connection layer. A convolutional neural network model with two-channel full connection layer is constructed to enhance the characteristic expression ability of the model. The proposed method can recognize 7 basic facial expressions with high accuracy, and has higher recognition accuracy compared with the single-channel facial expression recognition algorithm. Compared with other two-channel convolutional neural networks, it can achieve better recognition effect with simpler network structure.

Keywords: Double-channel Convolutional Neural Network; Facial Expression Recognition; Feature Fusion; Gradient Feature; LBP

1 Introduction

Facial expression is an important means of information transmission in interpersonal communication. Human facial expression carries the basic information of their current psychological state. In recent years, the investigation of facial expression recognition has attracted wide attention [2, 10]. Priya Saha [9] proved that face image collection is the best method to induce spontaneous emotion expression through emotion diagnosis and automatic analysis of video game player experience data.Facial expression recognition is one of the best ways to test any content, product or service that may trigger emotional arousal and facial reactions.Facial expression recognition has a reliable development space in many fields such as police investigation, online course monitoring, psychological research and intelligent traffic [4].

Facial expression recognition (FER) focuses on extracting facial expression features from facial images. The commonly used feature extraction methods are gradient features and texture features [15,16], and a trained classifier is used to recognize different facial expressions.However, traditional facial expression recognition technology [5] has interference factors such as illumination, pose and camouflage, etc. In order to better extract image features, Ojala [6] proposed Local Binary Patterns (LBP) to describe Local texture information of grayscale images.

However, the traditional LBP feature extraction method cannot fully display the feature information of the whole face, and cannot adapt to the interference brought by the expression and light change. Many scholars have proposed improved algorithms [3, 8, 14] to solve the above problems,but the feature extraction effect is still poor. Gradient domain image can represent the structure characteristics between adjacent pixels and can better describe the global information of human face. Therefore, the method of gradient feature fusion and LBP feature fusion can perform a more comprehensive representation of human face information.

2 Convolutional Neural Network Model

The basic structure of a typical CNN contain three different types of structure layer, convolution, pooling layer and the connection layer, as shown in Figure 1, through a series of convolution convolution layer pooling operation features extracted from the image, convolution pooling operation step by step, step by step a low-level feature mapping from the input data for high-level features, finally using high-level features for classification. The traditional working mode of convolutional neural network ignores the influence of low-level features on classification results, which results in that the constructed classifier is not the optimal classifier. In addition, with the deepening of network layers, the traditional neural network is prone to gradient dispersion problem, which leads to the training difficulty of the network.

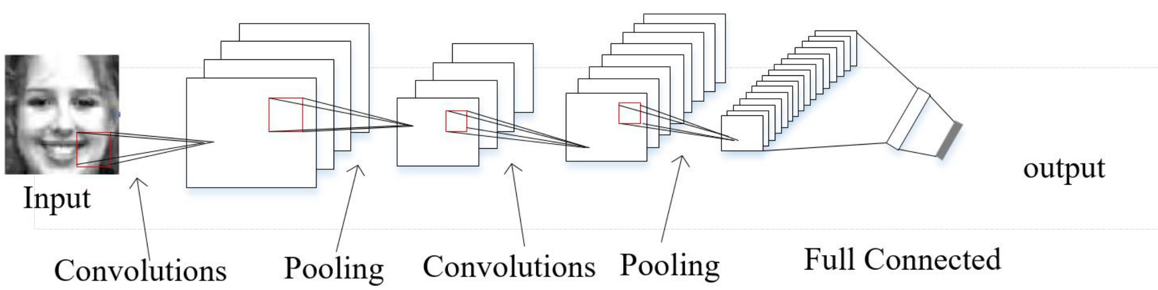

Figure 1: Basic structure of convolutional neural network

3 A Two-channel Convolutional Neural Network Model

In this paper, a dual-channel convolutional neural network model is designed. The structure of the feature extraction network is shown in Figure 2. Each channel consists of 1 input layer, 3 convolutional layers and 3 pooling layers.In the first convolutional layer, the size of the convolution kernel parameter is set as 5×5 to fully extract the image features. In view of the larger the number of layers in the convolutional neural network, the larger the visual field, the size of the convolution kernel of the remaining convolutional layers is set as 3×3. At the same time, the pooling size of all pooling layers in the network is set as 2×2, so as to maintain the invariability of the extracted face features in geometric features and reduce the image dimension. Combined with Maxout activation function and Dropout technology, the internal structure of the network was optimized, and the traditional single-channel full connection layer was improved to construct a DCNN model with double-channel full connection layer.

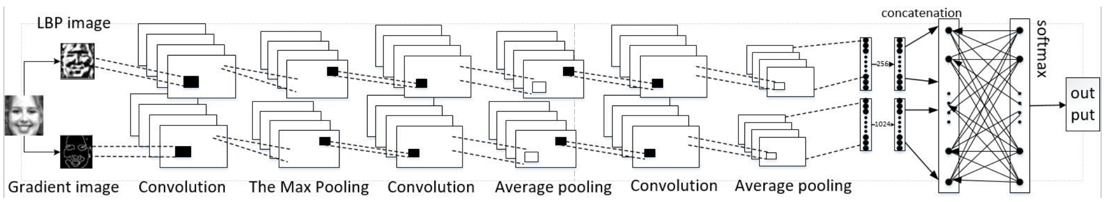

Figure 2: Two-channel convolutional neural network model

4 Experiments and Analysis

The deep learning tool used in this paper is Tensor Flow. All the experiments are carried out on Windows, where the computer hardware is configured as Intel Core I7 CPU, 16G memory, NVIDIA 960 graphics card and 2G video memory. The whole experiment process is implemented in Python language.

4.1 Database

The databases used in this chapter are JAFFE database and FER2013 database. Both are representative databases of spontaneous facial expressions, including variations in age, race, gender, posture, background, lighting and occlusion. It contains seven expressions: happy, sad, disgusted, afraid, surprised, angry and neutral. Figure 3 is a schematic diagram of 7 sample expressions in JAFFE database.Figure 4 is the schematic diagram of 7 expressions in FER2013 database.

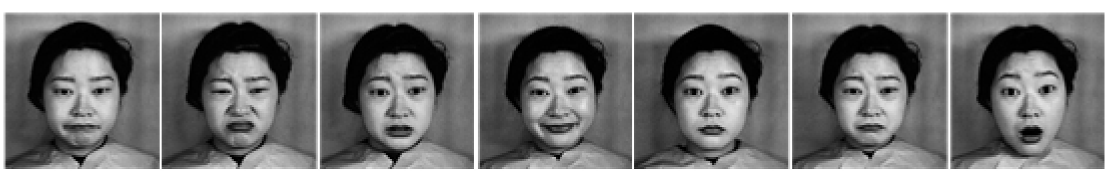

Figure 3: Some images of JAFFE emojis

4.2 Experimental results and analysis

The proposed model was used to carry out experiments on FER2013 database. ReLU activation function was used in convolutional layer and full connection layer, cross entropy loss function was used in training model, initial learning rate was 10E-4, and Dropout size of convolutional layer and full connection layer was 0.9. In the training model, experiments were carried out on three groups of double-channel full connection layers with different Numbers of neurons. The three groups of double-channel parameters were:

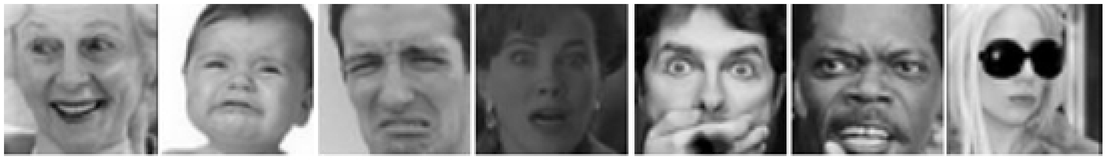

Figure 4: Some images of FER2013 emojis

1) Channel 1: 1024-1024, channel 2: 256-256;
2) Channel 1: 2048-1024, channel 2: 1024-256;
3) Channel 1: 4096-4096 and channel 2: 256-256 were used for one experiment respectively with the above three groups of parameters.

The results are shown in Table 1.

Table 1: The accuracy of different channel parameters of the proposed model

	Channel 1	**Channel 2**	**Accuracy (Percent)**
Group 1	1024-1024	256-256	68.34
Group 2	2048-1024	1024-256	58.54
Group 3	4096-4096	256-256	61.48

Table 2 and Table 3 show the confusion matrix of different expressions on FER2013 and JAFEE databases by using the improved model. It can be seen that the accuracy of the two models in predicting happy labels is relatively high, which indicates that happy expression features are easier to be recognized than other expression features. Compared with the traditional CNN model and the proposed model, the improved model can improve the recognition accuracy of most emoji tags.

It can be seen from Table 4 and Table 5 that the proposed method in this paper is superior to the traditional feature extraction method such as DCNN [13], Multi-scale CNN [12], VGG [11] on FER2013 and LBP [17], Gabor [1], and Gabor+SVM [7] on JAFEE.

Table 2: The obfuscation matrix of the proposed model on FER2013 database

	Angry	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**	**Neutral**
Angry	0.67	0.13	0.07	0.01	0.10	0.02	0.04
Disgust	0.11	0.72	0.09	0.02	0.06	0.03	0.01
Fear	0.04	0.01	0.61	0.04	0.21	0.11	0.01
Happy	0.01	0.00	0.01	0.93	0.04	0.00	0.00
Sad	0.09	0.05	0.15	0.03	0.63	0.04	0.02
Surprise	0.01	0.03	0.02	0.06	0.01	0.86	0.00
Neutral	0.03	0.00	0.10	0.05	0.03	0.01	0.75

Table 3: The obfuscation matrix of the proposed model on JAFEE database

	Angry	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**	**Neutral**
Angry	0.91	0.05	0.01	0.00	0.02	0.01	0.00
Disgust	0.07	0.83	0.02	0.00	0.03	0.02	0.00
Fear	0.01	0.04	0.84	0.02	0.05	0.03	0.01
Happy	0.01	0.01	0.00	0.95	0.01	0.03	0.00
Sad	0.03	0.03	0.11	0.02	0.81	0.01	0.00
Surprise	0.00	0.02	0.04	0.05	0.01	0.86	0.00
Neutral	0.02	0.01	0.03	0.01	0.01	0.02	0.87

Table 4: Comparison of recognition accuracy of different CNN methods in FER2013

Methods	**Accuracy (Percent)**
DCNN [13]	72
Multi-scale CNN [12]	72.82
VGG [11]	72.72
Proposed Method	76.86

Table 5: Comparison of recognition accuracy of different CNN methods in JAFEE

Methods	**Accuracy (Percent)**
LBP [17]	81.6
Gabor [1]	80.1
Gabor+SVM [7]	86.9
Proposed Method	88.75

5 Conclusion

To solve the problem of insufficient feature extraction and low recognition rate, an improved two-channel convolutional neural network is proposed to recognize 7 facial expressions. Using LBP image and gradient image of human face, both global and detailed information of human face are considered.Experimental results show that the recognition rate of the feature extraction method in FER2013 and JAFEE data sets is significantly improved. At the same time, the expression recognition performance of the proposed method is better than that of the single-channel CNN expression recognition algorithm. At present, There is no universal structure for CNN to solve a variety of problems, and different problems need to be solved by designing different structures, which brings certain difficulties to the popularization of CNN. Therefore, how to design a CNN structure so that it can solve a variety of problems remains to be further studied.

References

[1] P. Z. Chen, C. Hu, X. Y. Chen, "Facial expression recognition with Gabor, LBP and LPQ features," *Measurement and Control Technology*, vol. 37, no. 08, pp. 16-20, 2016.

[2] C. A. Corneanu, M. O. Simon, J. F. Cohn, *et al.*, "Survey on RGB, 3D, thermal, and multimodal approaches for fFacial expression recognition: History, trends, and affect-related applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1548-1568, 2016.

[3] Z. H. Guo, L. Zhang, D. Zhang, *et al.*, "Rotation invariant texture classification using adaptive LBP with directional statistical features," in *17th IEEE International Conference on Image Processing*, pp. 285-288, 2010.

[4] Y. Huang, F. Chen, S. Lv, X. Wang, "Facial expression recognition: A survey," *Symmetry*, vol. 11, no. 10, 2019.

[5] T. Jabid, M. H. Kabir, O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI Journal*, vol. 32,no. 5, pp. 784-794, 2010.

[6] T. Ojala, M. Pietikinen, D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51-59, 1996.

[7] L. V. Praseeda, M. Sasikumar, "Analysis of facial expression using Gabor and SVM," *International Journal of Recent Trends in Engineering*, vol. 1, no. 2, pp. 47-50, 2009.

[8] M. M. Qi, Y. Xiang, "Face recognition based on MB-LBP and improved LFDA," *Computer Science*, vol. 39, no. 6, pp. 266-269, 2012.

[9] P. Saha, D. Bhattacharjee, B. K. De, *et al.*, "A survey on image acquisition protocols for non-posed facial expression recognition systems," *Multimedia Tools and Applications*, vol. 78, no. 16, pp. 23329-23368, 2019.

[10] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1113-1133, 2015.

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhou, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)*, 2015.

[12] J. Wang, C. Yuan, "Facial expression recognition with multi-scale convolution neural network," in *17th Pacific-Rim Conference on Advances in Multimedia Information Processing*, pp. 376—385, 2016.

[13] Z. Yu, "Image based static facial expression recognition with multiple deep network learning," in *ACM on International Conference on Multimodal Interaction*, 2015.

[14] B. H. Yuan, H. Wang, and M. W. Ren, "Face recognition based on completed local binary pattern," *Application Research of computers*, vol. 29, no. 4, pp. 1557-1559, 2012.

[15] S. Zhang, J. Gong, D. Chen, *et al.*, "Sparsity-motivated multi-scale histograms of oriented gradients feature for SRC," in *2017 IEEE International Conference on Unmanned Systems(ICUS'17)*, pp. 383-393, 2017.

[16] G. Zhao, M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 915-928, 2007.

[17] X. Zhao, S. Zhang, J. Liu, S. Yin, "Facial expression recognition based on local binary patterns and kernel discriminant isomap," *Sensors*, vol. 11, no. 12, pp. 9573-9588, 2011.

Biography

Tingting Gao biography. She received the B.Eng. degree from Shenyang Normal University, Shenyang , Liaoning province, China in 2018. Now, she is studying for Master degree in Software College, Shenyang Normal University. Her research interests include Multimedia Security, Network Security, image processing and Data Mining. Email:841584960@qq.com.

Hang Li biography. He is a full professor in Software College, Shenyang Normal University. He received his B.S. and M.S. degrees from Northeastern University. He is also a master's supervisor. He has research interests in wireless networks, mobile computing, cloud computing, social networks, network security and quantum cryptography. Professor Li had published more than 30 international journal papers (SCI or EI journals) on the above research fields. Email: lihangsoft@163.com.

Shoulin Yin biography. He received the B.Eng. and M.Eng. degree from Shenyang Normal University, Shenyang , Liaoning province, China in 2016 and 2013 respectively. Now, he is a doctor in Harbin Institute of Technology. His research interests include Multimedia Security, Network Security, Filter Algorithm, image processing and Data Mining. Email:yslinhit@163.com.