Advances in Artificial Intelligence, Big Data and Algorithms
G. Grigoras and P. Lorenz (Eds.)

© 2023 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA230873

Facial Expression Recognition in Classroom Environment Based on Attention Mechanism

Yue WU[1]
College of Electrical and Information Engineering, Jilin Engineering Normal University

Abstract: In view of the lack of facial expression data set in the classroom environment, the classroom expression data set was constructed, including the acquisition and preprocessing of students' face pictures, the selection of students' emotional categories in the classroom environment and the labeling of pictures. Based on the Resnet50 network model, a network structure with attention module is proposed, so that it can focus on the feature parts that clearly represent the target emotion in facial images, so as to enhance the accuracy of facial emotion recognition. In order to verify the effect of the model presented in this paper, training tests were carried out on the common data set of expression Fer2013 and the classroom data set constructed in this article. The results show that the structural model presented in this paper has better recognition effect and can effectively enhance the accuracy of expression recognition.

Keywords: Resnet-50; SE; lightweight ; facial expression recognition

1. Introduction

With the rapid development of the field of artificial intelligence, education has also stepped into the era of intelligence, and the intelligent identification of students' learning state in the classroom has become more and more important. In the traditional classroom, due to the large number of students, teachers cannot ensure timely adjustment of the course progress, Mehrabian, a psychologist, pointed out that, 55% of emotional information is expressed through facial expressions. Therefore, in the classroom environment, teachers can judge students' emotional changes in real time through facial expressions, make timely teaching adjustments. A classroom expression database is established to solve the problem that there are few existing data sets of classroom expression recognition. In order to solve the defect of insufficient extraction ability and inadequate extraction of key features in complex images, a new recognition model with attention mechanism is proposed.

[1] Corresponding Author: Yue WU, College of Electrical and Information Engineering, Jilin Engineering Normal University; e-mail: 28841898@qq.com

2. Classroom Student Expression Recognition

Classroom emotional understanding can be divided into two ways based on traditional manual characteristics and deep learning.

2.1 Traditional Classroom Emotional Understanding Methods

Traditional emotion recognition research generally uses low-level manual features. Commonly used feature extraction methods include principal component analysis (PCA) [1], local binary pattern method (LBP) [2] and Gabor transform method [3]. Commonly used classification methods based on machine learning include SVM, K-nearest neighbor algorithm, etc. The traditional method can not obtain the deep features of the image, and has some problems such as large amount of computation and easy loss of local details, so it can not be applied in classroom emotion analysis.

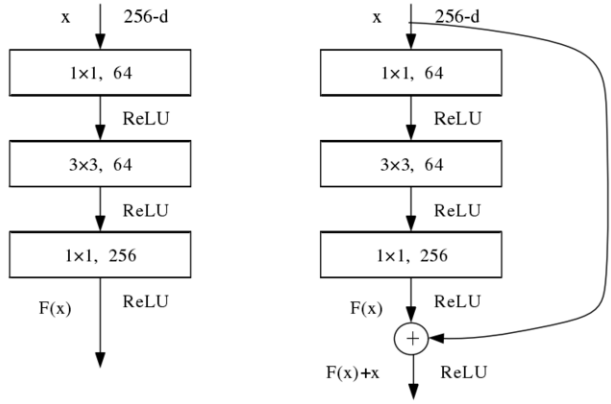
2.2 Expression Recognition of Classroom Students Based on Deep Learning

Expression recognition algorithm can do a good classification of emotions With the development of deep learning. At present, mainstream expression recognition models include VGGNet[4], GoogleNet[5],AlexNet[6]. However, with the deepening of the network layer, the phenomenon of gradient explosion will become more and more serious. In order to solve this problem, He Kaiming et al. proposed deep residual network Res Net[7]. The residual module is added to the network to alleviate the problem of network degradation when the network layers are too deep. Therefore, this paper chooses Res Net-50 as the basic network model and combines attention mechanism to identify students' expressions in the classroom environment.

3. Network Model Structure

3.1 Residual Structure

Figure. 1 is a comparison diagram of the ordinary directly connected structure and the residual block structure. Each of the three convolution layers is followed by a ReLU activation function layer and a batch normalization layer (BN layer), whether it is an ordinary directly connected structure or a residual block structure.Different from the directly connected structure, as shown in Figure 1 (b), the residual block also introduces a bypass branch line, which directly transmits the input feature to the following layer, so that the latter layer can directly learn the input feature and learn more complete information, so as to alleviate the phenomenon of network degradation.Instead of directly fitting a direct complete map between input and output like previous convolutional neural networks, residual networks learn the difference between output and input.Due to the particularity of the residual structure, when the input is differentiated by loss, the derivative term will be decomposed into two. In the process of backpropagation, the gradient disappearance or gradient explosion will not be caused by the deepening of the residual network, and the network iteration is more stable.



(a) Common directly connected structure (b) Residual block structure

Figure 1. Contrast diagram of the common directly connected structure and the residual block structure

3.2 Squeeze & Excitation Attention

SEnet module [8] is a feature information channel attention module that can effectively enhance channel dimension. SEnet takes care to get the output of the convolutional block and converts each channel to a single value via a global average pool; This process is called "squeezing." After increasing nonlinearity through the full connection layer and ReLU, the output channel ratio decreases. These features pass through fully connected layers followed by an S-shaped function to achieve smooth gating operations. The convolutional block feature map is weighted based on the output of the side network, which is called "excitation". This process can be summarized as:

$$f_s = \sigma(FC(\mathrm{Re}LU(FC(f_g)))) \tag{1}$$

Fc is the fully connected layer, fg is the average global pooling layer, and is the sigmoid operation. The basic structure of the SEnet module is shown in Figure 2.

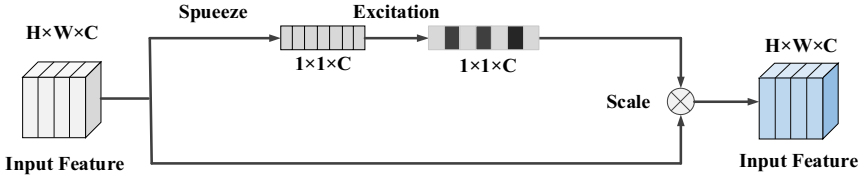


Figure 2. Structure diagram of the SE module

3.3 The Embedding of Attention Modules

The introduction of the attention mechanism module can effectively improve the performance of the model, but the result will not be improved if arbitrarily added into the network. Therefore, it is necessary to find the most suitable insertion position and determine the best embedding position of the module. Experimental analysis was conducted on the addition positions of the three modules shown in Figure 3.

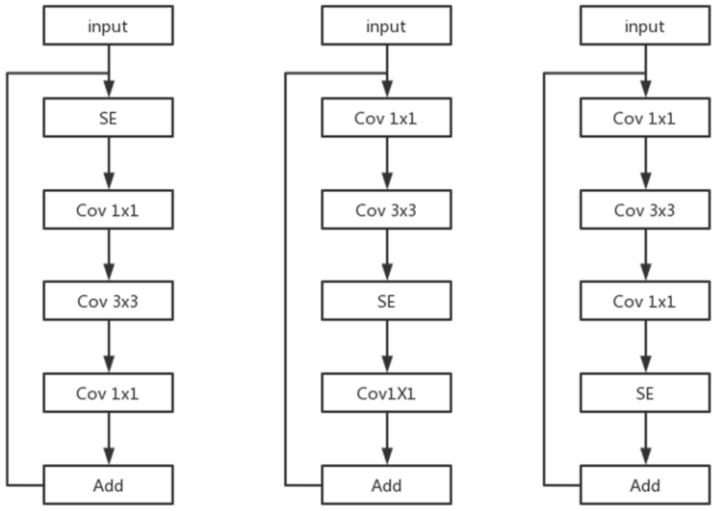


Figure 3. The SE module is embedded in different positions in Resnet50

4. Experimental Results and Analysis

4.1 Data Set

In this paper, public facial expression data set FER2013 and self-built classroom expression data set were used for verification.

Fer-2013 is a dataset provided for the Kaggl Facial Expression Recognition competition. The dataset contained seven categories: anger (4953), disgust (547), fear (5121), happiness (8989), sadness (6077), surprise (4002), and neutrality (6198). There are obvious differences in facial posture, age, intensity of expression, and skin color of similar expressions, and many faces are blocked by objects such as glasses, hats, and hands, which are more consistent with facial expressions in real scenes.

Since there is no public classroom expression database on the Internet, this paper constructs a set of classroom expression database by collecting data from a real classroom in a university.,Including disdain, distraction, satisfaction , fatigue , confusion and concentration. This data set is mainly used for model validation.

4.2 Experimental Setting

The experiment was conducted on TensorFlow, an open source deep learning framework, and the platform was Anaconda3. On all data sets, the learning rate was initialized to 0. 1. A total of 30 Epoch experiments were performed.

Experiments were conducted in Fer2013 data set to explore the influence of different SE module locations on the accuracy. The experimental results are shown in Table 1. The results show that the SE module is located in front of the first 1*1 convolution in the network, the accuracy is the highest, reaching 71.38%.

Table 1. Influence of different SE modules on accuracy in Resnet-50 network

Network structure	Accuracy	parameters
RESNET50	70.96%	23,602,055
01	71.38%	23,494,637
02	70.66%	23,028,679
03	71.05%	20,161,607

4.3 Real-Time Test

In order to verify the generalization of the model, this paper uses self-made data set to verify the generalization of the model. Figure 4 shows the three expressions collected in the self-made data set, namely, distraction, satisfaction and focus (due to privacy concerns, the student images displayed were mosaicked in their key parts). Table 2 shows the test results of these expressions.

Table 2. Accuracy of model in distracted, satisfied and focused data

Network structure	Accuracy	parameters
RESNET50+SE	91.23%	23,081,943

(a) distracted

(b) satisfied

(c) focused

Figure 4. Driving emotions of distracted, satisfied, focused data

5. Conclusion

Aiming at the problem that the classroom facial expression recognition algorithm tends to ignore some important features with the number of network layers increase in the feature extraction stage, this paper proposes an introduction of attention mechanism module to enhance the feature extraction ability of the network. The network can enhance significant features, so as to reduce the rate of missing detection, and improve the network expression ability. The experimental results show that this model can reach high accuracy for classroom facial expression recognition.

Acknowledgments

This work is supported by the 13th Five-Year Plan Project of Education Science of Jilin Province. Project name: Analysis and Research on the Development Status of Artificial Intelligence Industry in Jilin Province and its Demand for Applied Talents (ZD20039)

References

[1] WOLD S.ESBENSEN K.GELADIP.Principal component analysis[J].Chemometrics and Intelligent Laboratory Systems,1987,2(1/2/3/):37-52.

[2] OJALA T,MAKNPAA T,PIETIKAINEN M,et al.Outex-new framework for empirical evaluation of texture analysis algorithms[C]//2002 International Conference on Pattern Recognition.IEEE,2002:701-706.

[3] ZHANG Z,MU X,GAO L.Recognizing facial expressions based on gabor filter selection[C]//2014th International Congress on Image and Signal Processing.IEEE,2014:1544-1548.

[4] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. ar Xiv preprint ar Xiv:1409.1556, 2014.

[5] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015:1-9.

[6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Image Net classification with deep convolutional neural networks[C]//Proceedings of the 2012 Advances in neural information processing systems. New York: ACM, 2012: 1097-1105.

[7] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
[8] HU J , SHEN L,SUN G.Squeeze-and-excitation networks[C].Proceeding of the IEEE Conference on Computer Vison and Pattern Recognition.2018:7132-7141.