2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE)

Facial Expression Recognition Method Based on Convolutional Neural Network and Data Enhancement

Lisha Yao [1,2]
[1.]School of Big Data and Artificial Intelligence, Anhui Xinhua University
Hefei, Anhui, China
[2.] College of Computing and Information Technologies,National University
Manila, Philippines
*jsjyaolisha@163.com

Kang Su
School of Big Data and Artificial Intelligence, Anhui Xinhua University
Hefei, Anhui, China
2309620371@qq.com

2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE) | 978-1-6654-0886-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCECE54139.2022.9712809

***Abstract*—The effect of direct data on original facial expression recognition is not good. In order to further improve the expression recognition effect of convolutional neural network, this paper proposes an expression recognition method based on convolutional neural network and data enhancement. This paper improves the existing Alexnet network and designs a suitable network structure. At the same time, data enhancement was carried out on FER2013 data set to improve the effect of facial expression recognition. Experiments prove the effectiveness of the proposed method and verify that the proposed algorithm improves the accuracy of facial expression recognition to a certain extent.**

Keywords-facial expression recognition; CNN; data enhancement

I. INTRODUCTION

With the continuous development of artificial intelligence, the demand for human-computer interaction technology also increases. Facial expression is a basic way of expressing human emotions and the main way of interpersonal communication[1]-[2]. Numerous expressions complement language well and can better express one's thoughts. People judge each other's inner activities by expressing their emotions. In addition, the classification of facial expressions is an important research direction of human emotion exploration and cognitive psychology.

Convolutional neural network(CNN) is one of artificial intelligence learning frameworks[3]-[4]. The network is robust to illumination, translation and rotation. Neurons at all layers of CNN share weights and adopt the method of incomplete connection, which greatly reduces the parameters of network training and reduces the complexity of the network. Compared with neural interconnection network, CNN has its unique characteristics.

In 1971, American researchers Aikman and Friesen [5] laid a solid foundation for the recognition of current human facial expression features. Ekman classifies six basic expressions made by human beings[6]: happiness, anger, surprise, fear, disgust and sadness, and demarcates which categories are the targets of discrimination recognition. Then, he established a facial action Coding system (FACS), according to which researchers can divide relevant facial action units to describe facial actions and detect subtle facial expressions through the relationship between facial actions and expressions. In 2012, Krizhevsky et al. achieved unexpected achievements by using Alexnet[7]-[8] network in ILSVRC-2012, which could not be matched by traditional methods.

To sum up, in order to further improve the identification effect of the network and solve the over-fitting problem caused by the small data set. This paper is optimized and improved based on Alexnet and combined with data enhancement to improve the accuracy of facial expression recognition.

II. INTRODUCTION TO CNN STRUCTURE

A. Convolution Layer

In CNN, each convolution layer is constituted by one or more convolution units. The main function of the back propagation algorithm is to improve the parameter variables of the convolution unit at the network layer and then optimize the algorithm. Convolution is the main method used to obtain various features of the incoming data. In the process of obtaining image information at the first layer, only some edge features may be extracted. In the deeper network layer, features extracted earlier can be reprocessed to further obtain more accurate and representative features.

The parameter variables of the convolution layer include a series of convolution kernels (filters). The convolution kernels move in parallel on the image and do inner product and sum with the pixels in the moved region, so as to obtain the characteristic parameters of each specific position. The calculation formula of convolution kernel at different depths is:

$$a_{d,i,j} = f\left(\sum_{d=0}^{D-1}\sum_{m=0}^{F-1}\sum_{m=0}^{F-1} w_{d,m,n}x_{d,i+m,j+n} + w_b\right) \quad (1)$$

D is the depth. F is the size of filter. $w_{d,m,n}$ is the weight variable of m rows and n columns of the convolution kernel at the d-layer. $a_{d,i,j}$ is the pixel of image layer d, row i and column j. $x_{d,i+m,j+n}$ *represents pixels in d,* $i+m$ *rows and* $j+n$ *columns of the image layer.* w_b *is the offset value.*

978-1-6654-0886-8/22/$31.00 ©2022 IEEE

B. Pooling Layer

In order to better reduce the parameters of the network and improve the training effect of the model, pooling/sampling is used in the experiment. The complexity of calculation is reduced by reducing the size of the feature graph. One is to obtain key features through feature compression. Pooling/sampling is divided into average pooling and maximum pooling. The pooling operation mainly changes the size, other conditions do not change.

When people express emotions, they will show the categories of expressions through the changes of eyes, nose and mouth. Through maximum pooling, the neurons that respond most strongly to expressions and actions can be reserved. Take the incoming data image of size $W_i \times H_i$, width is W_i, height is H_i, the height and width of convolution kernel are F, S is the distance moved in each step of the filter, and finally the size of the output image through calculation is:

$$W_o \times H_O = \left(\frac{W_i - F}{S} + 1\right) \times \left(\frac{H_i - F}{S} + 1\right) \quad (2)$$

C. Fully Connected Layers

Fully connected layers(FC) whose main task is to classify categories. Fully connected layers has another function, which is to pass the acquired features into the labeled sample space in the way of mapping. The so-called full connection refers to the connection of a neuron and all the neurons at the upper level and the lower level of its layer. The power of the full connection layer is that it can classify and divide the shallow local information extracted from the feature extraction layer. In order to improve the experimental effect, Relu was mainly used as the activation function in the experiment. After the final processing, the output parameters are passed to a classification function, which is usually classified by Logistics Regression. Softmax layer plays a key role in classification. The loss function plays a key role in the experimental results. Generally speaking, most of the algorithms used by CNN are back-propagation algorithms.

D. Softmax Layer

The reason for using Softmax is that it maps results to a range of 0 to 1, where the sum of all results equals 1. If this function is used for multiple classification experiments, it will return the percentages of each category, and the percentages of this category will be much higher than the percentages of other categories. The formula of Softmax is:

$$\mathrm{P}(i) = \frac{\exp\left(\theta_i^T x\right)}{\sum_{k=1}^{K} \exp\left(\theta_k^T k\right)} \quad (3)$$

You can observe that all the values add up to exactly 1, and all the outputs map to the region from 0 to 1, which is what we call probability problems. $\theta_i^T x$ has multiple inputs, and the original purpose of training is to approximate the best θ^T.

III. Facial Expression Recognition Method Based on Convolutional Neural Network and Data Enhancement

A. Overall Thinking

In this paper, Alexnet is improved to make it suitable for face expression experiment. At the same time, in order to improve the recognition effect, the data is enhanced.

The steps of the method proposed in this paper are as follows:

- Step 1: Image preprocessing, including data enhancement to expand the data set.
- Step 2: Train the network through improved Alexnet.
- Step 3: Use test sets to test network performance to complete identification.

B. Alexnet Network Model

Alexnet is heavily used in computer vision, a network that Hinton's Google team (Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton) won in 2010. Its network structure is shown in Figure 1:

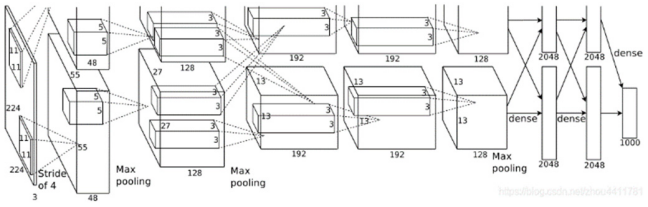

Figure 1. Alexnet structure.

The ReLu activation function used by Alexnet is compared with Tanh mainly because the calculation speed of ReLu activation function is relatively fast in gradient descent calculation.

Data enhancement plays an important role in the field of vision, mainly in order to reduce the probability of over-fitting. The two methods used in this network are horizontal rotation and changing the number of RGB.

This network uses two optimization methods, ReLu and Dropout, to improve the computational speed and reduce the probability of over-fitting. Although these methods are only summarized by the author, he is the first one who has actually applied them to practical experiments, which also provides inspiration for other researchers.

C. Improved Alexnet Network Model

In this paper, Alexnet is improved to make it suitable for face expression experiment. At the same time, in order to improve the recognition effect, the data is enhanced.

The network used in this experiment is summarized on Alexnet. The number of Convolution layer, Pooling layer and connection layer in the current network structure layer is 3. The network structure is shown in Figure 2:

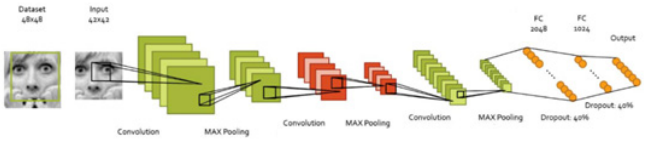

Figure 2. Improved Alexnet Network structure

The network details are shown in Table I:

TABLE I. MODEL SCHEMA TABLE

type	*kernel_size*	*kernel_num*	*pad*	*output*	*dropout*
Data				42×42×1	
Convolution	5 × 5	32	2	42×42×32	
Pooling	3 × 3			21×21×32	
Convolution	4 × 4	32	1	20×20×32	
Pooling	3 × 3			10×10×32	
Convolution	5 × 5	64	2	10×10×64	
Pooling	3 × 3			5×5×64	
InnerProduct				1×1×2048	0.4
InnerProduct				1×1×1024	0.4
InnerProduct				1×1×7	

IV. EXPERIMENTAL ANALYSIS

A. Data preprocessing

According to the experience of the previous experiment, neural networks are more prone to over fitting phenomenon, in order to avoid the occurrence of this phenomenon, typically by enhance the way to improve the adaptability of the network data: one is the expansion of data sets to improve the experiment ability of the model, the second is accurate training data can improve computing power and accuracy of the model. Commonly used methods are as follows: parallel movement, change size, transpose operation, color change, visual change, partial cover.

The experiment is divided into three times. Experiment 1 adopts the original FER2013 data set, which has a total of 35887 face images, including 28709 training, 3589 verification and 3589 test images respectively. In experiment 2, fer2013 data set was obtained after data enhancement by some operations such as flipping, translation and rotation. The enhanced database contained 84021 face images, including 76843 training, 3589 verification and 3589 test images respectively. In experiment 3, the size of images in FER2013 data set was changed to 256×256 after processing, and the Alexnet neural network model was used for training.

(a) (b) (c) (d)

(e) (f) (g)

Figure 3. Sample Diagram of Data Enhancement.

Figure 3 shows that (a) is the original data in FER2013 dataset, and (b) is the enhanced image obtained by flipping. (f), (g) enhanced data obtained by translation of (b), (c) images obtained by inversion of (a), (d) is the inversion of (c). It is through the original data of the flip, translation and other methods to fill the data, and then increase the number of original data to meet the requirements of the experiment.

At the same time, the experiment changes the size of the image to adapt to the size requirements of different neural networks on the input data, so as to compare the influence of different neural networks on the experiment.

As shown in Figure 4, the accuracy of the three experiments keeps improving. Meanwhile, it can be observed that the accuracy of experiment 3 is higher than that of Experiment 1 and Experiment 2, indicating that a deeper network structure can extract feature information more effectively and is more beneficial to the experiment.

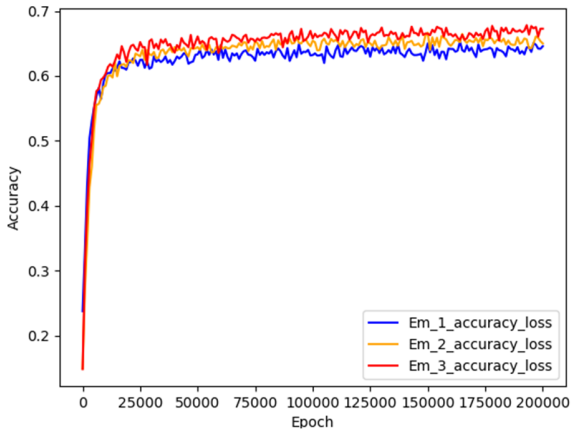

Figure 4. Comparison Chart of Accuracy in Experiment 1, 2 and 3.

As shown in Figure 5, in the three experiments, it can be seen that the average test_loss value using data enhancement is the lowest, while the average value of the experiment using deeper network structure is the highest. Therefore, the over-fitting degree of experiment 3 is higher, while the over-fitting degree of experiment using data enhancement is significantly reduced.

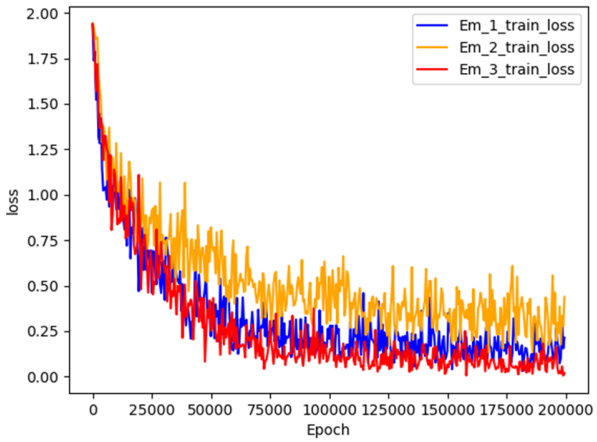

Figure 5. Comparison Chart of Train_loss in Experiment 1, 2 and 3.

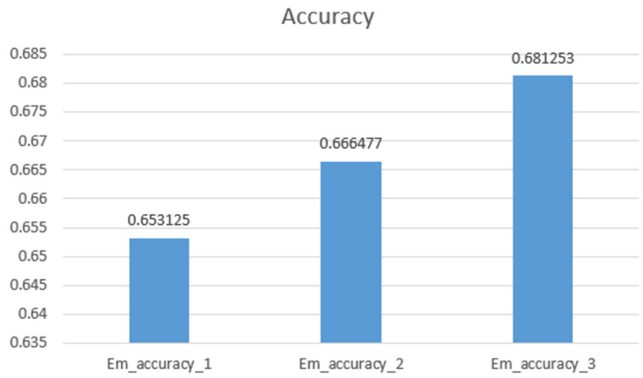

Figure 6. Accuracy Histogram of Experiment 1, 2 and 3.

It can be seen from Figure 6 that the accuracy of experiment 2 has been improved by nearly 1.34% compared with experiment 1 after data enhancement. In experiment 3, the deeper neural network was used to improve the accuracy by nearly 2.81% compared with experiment 1. The accuracy values for Experiments 1, 2, and 3 are the maximum values in the dataset extracted by PANDAS.Based on the above summary, it is concluded that the expansion of data set and deepening of network depth can improve the accuracy of experimental model, but under the condition of unchanged data set, the phenomenon of over-fitting is more likely to occur in experiment 3.

TABLE II. EXPERIMENTAL DATA

	number of data sets	*Accuracy*	*Number of training*
Fer2013_1	35887	0.653125	200000
Fer2013_2	84021	0.666477	200000
Fer2013_3	35887	0.681253	200000

In Table II, Fer2013_1, Fer2013_2 and Fer2013_3 are the original data, enhanced data and size change data respectively. It can be seen from Table II that the increase of data sets and network depth can improve the value of Accuracy.

V. CONCLUSION

This paper proposes an image recognition method based on CNN and data enhancement. It continuously optimizes Alexnet to improve network identification ability. At the same time, data enhancement on FER2013 data set verifies the efficiency and authenticity of the proposed method in facial expression feature classification. In this paper, static pictures were used in the experiment instead of data pictures that changed at any time, which also resulted in the loss of some information features and had a great impact on the experiment. In the future, experiments will be carried out on the way of dynamic video frame extraction.

ACKNOWLEDGMENT

Special thanks to the following funds for their support: Key Research Project of Natural Science in Universities of Anhui Province(No.KJ2020A0782);University-level Quality Engineering Demonstration Experiment and Training Center "Big Data Comprehensive Experiment and Training Center" (No. 2020 sysxx01); 2020 Anhui Provincial College Student Innovation Plan Project (No. 202012216083).

REFERENCES

[1] Liu Q M, Xin Y Y. Face Expression Recognition Based on End-to-End Low-quality Face Images[J]. Journal of Chinese Computer Systems, 2020, 041(003):668-672.

[2] Li T T,Hu Y L,Wei F L. Improved Facial Expression Recognition Algorithm Based on GAN and Application[J]. Journal of Jilin University(Science Edition), 2020,058(003):605-610.

[3] Lopes A T, Aguiar E D, Souza A, et al. Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order[J]. Pattern Recognition, 2017, 61:610-628.

[4] Ding M D,Li L. CNN and HOG Dual-path Feature Fusion for Face Expression Recognition[J]. Information and Control, 2020, 49(1):47-54.

[5] Ekman P,Hager J C,Friesen W V.The Symmetry if Emotional and Deliberate Facial Actions[J]. Psychophysiology,2010,18(2):101-106.

[6] He Z C,Zhao L Z,Chen C. Convolution Neural Network with Multi-Resolution Feature Fusion for Facial Expression Recognition[J]. Laser & Optoelectronics Progress,2018,55(07):370-375.

[7] Han X , Zhong Y , Cao L , et al. Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification[J]. Remote Sensing, 2017, 9(8):848.

[8] Li A H , Luo Y , He Y H , et al. Fault diagnosis method of rare earth extraction production line based on wavelet packet and alexnet transfer learning[J]. Journal of Physics Conference Series, 2021, 1820(1):012102.