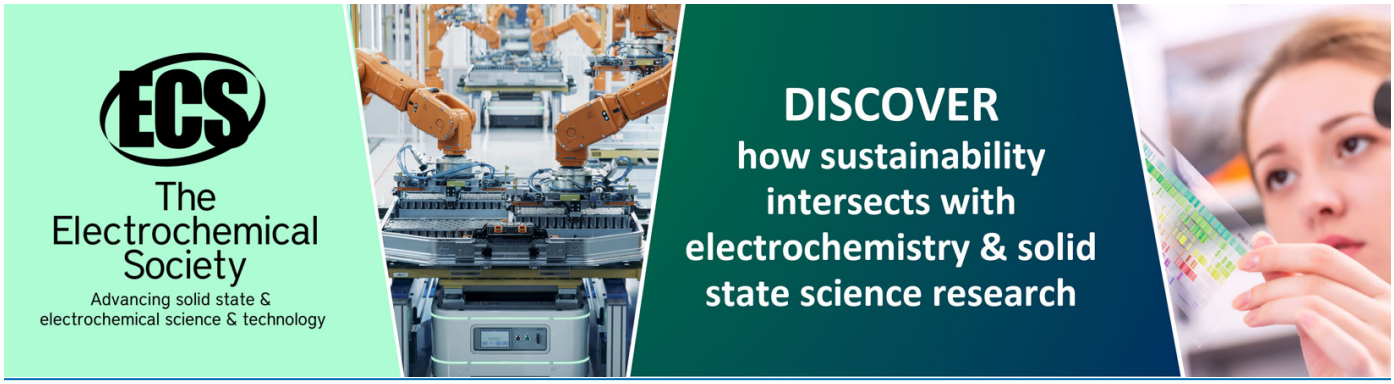
Facial expression recognition based on CNN



Facial expression recognition based on CNN

Mingjie Wang[1], Pengcheng Tan[2], Xin Zhang[3], Yu Kang[4], Canguo Jin[5], Jianying Cao[1*]

[1] School of information engineering, Longdong university, Qingyang, Gansu 745000, China

[2] The School of Computer Science and Technology, Anhui University of Technology, Maanshan, Anhui 243032, China

[3] College of Computer Science and Technology, St. Petersburg Polytechnic University, Saint Petersburg, SP 190121, Russia

[4] College of resources and environment, Chengdu University of Information Technology, Chengdu, Sichuan 610225, China

[5] School of Information Science & Engineering, Lanzhou University, Lanzhou, Gansu 730000, China

Corresponding author's e-mail: cao_jianying@163.com

Abstract. Facial expression recognition usually can be defined one of the important research in the field of AI and pattern recognition. Facial expression contains rich invisible information, which can help to understand human emotions and intentions, which has great research value. Aiming at solve problems that usually happened such as low recognition accuracy and weak generalization ability of traditional facial expression recognition methods, this paper introduces the theory of the convolution neural network and establishes the convolution neural network model to realize facial expression recognition. Using the crawler to crawl the network image data and using the Viola-Jones algorithm to detect the original data set, after screening, the face expression database is finally established. The convolution neural network model is applied to face expression recognition of the database. We can find that this method can effectively recognize facial expression, and provides a new way to solve this problem.

1. Introduction

We all know that facial expression recognition technology, as an very important direction of emotion computing research, is an important part of human-computer interaction, and has a wide range of applications in medicine, education, business marketing and other fields. Mehrabian, a famous American psychologist, proposed that in the daily communication of human beings, the information transmitted by language and voice accounted for 7% and 38% of the total information respectively, while the information transmitted by facial expression accounted for 55%. American psychologists proposed that there are six basic human expressions through a large number of experiments. The recognition method based on feature is the key to the expression recognition of classifier. The traditional classification method needs to extract feature manually for classification. The quality of feature selection directly determines the accuracy of recognition, while the feature selection needs some professional knowledge, and the recognition rate is low, time-consuming and laborious. In recent years, deep learning, as a new study direction of machine learning, has been widely studied by scientists. Deep learning has

a significant improvement in timeliness and accuracy. Convolutional neural network (CNN) is an algorithm of deep learning. Lecun first proposed its idea in 1989, and in 1998 proposed the application of this algorithm to handwritten digit recognition[1-3]. In 2012, Alex Krizhevsky won the Imagenet 2012 competition with CNN. CNN can input image directly and get the final classification result without data preprocessing. By building a neural network model with a certain depth and combining nonlinear operations such as convolution and pooling, we can realize two important functions of imitating the hierarchical processing of human brain and local perception of visual nerve. It has been proved that the network has achieved good results in face recognition, speech recognition, vehicle detection and target tracking[4-7].

2. Facial expression classification algorithm based on CNN

Because of computer computing great performance, deep learning has been widely used in many fields. The deep neural network based on CNNk is applied to the problem of expression classification in the paper. In each convolution layer, the upper expression features are convoluted by a learnable convolution kernel, and the higher dimensional expression features are output to the lower convolution layer. The calculation formula of expression feature graph is as follows

$$G_i = f(G_i * W_i + B_i) \tag{1}$$

In formula (1), G_i is the output of i-th layer neuron, G_{i-1} is the input of i-1-th layer I neuron; W_i is the weight vector of convolution core of i-th neuron, B_i is the offset vector, and f is the activation function. To some extent, the pool layer can keep the feature scale and reduce the feature map dimension. The pooling formula is:

$$G_i = Pooling(G_{i-1}) \tag{2}$$

Where GI is the lower sampling layer and pooling is the pooling function. As an important part of neural network, activation function can add nonlinear factors to better simulate the structure of human neural network, and can retain and map features.

The full connection layer connects the neurons in the upper layer and each neuron in the layer to realize the purpose of synthesizing the extracted features, so it is also called multi-layer perceptron, and its calculation formula is as follows

$$F(x) = f(x * W + B) \tag{3}$$

Where, $F(x)$ is the fully connected layer, f usually called the activation function, W usually called the weight vector, and B can be called the offset vector. SoftMax function is usually used to solve multi classification problems. In this paper, SoftMax function is used to map the output of multiple neurons to a value from 0 to 1, indicating that the probability of expression image x belonging to category j is

$$P\{y^{(i)} = j \mid x^{(i)}, \theta\} = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{i=1}^{k} e^{\theta_j^T x^{(i)}}} \tag{4}$$

In formula (3), $P\{y^{(i)} = j \mid x^{(i)}, \theta\}$ is the probability of expression image x corresponding to each label category, and θ is defined as parameter to be fitted. The CNN structure is shown in Table 1.

Table 1. Network structure

Type	Input size	Kernel size / step	Output size
Convolution	48×48	5×5/1	44×44
Convolution	44×44	3×3/1	42×42
Pooling	42×42	2×2/1	21×21
Convolution	21×21	3×3/1	19×19
Convolution	19×19	3×3/1	17×17
Convolution	17×17	3×3/1	15×15
Pooling	15×15	2×2/1	7×7
SoftMax	Classified output		

3. Face detection algorithm

The key point of face detection(FD) we defined that is find out the location of the human face exist in the image. Whether face recognition or expression recognition, FD is a necessary step. In face expression recognition, face feature extraction is required. If the face is not located, there will be other redundant information when the image is input to the network or feature extraction is carried out, which will seriously affect the recognition effect, so face detection is very important. The FD algorithm we used is the Viola-Jones algorithm. This algorithm uses AdaBoost and Haar face detection technology to treat Haar features as a weak classifier. Then, in this algorithm, multiple weak classifiers are combined into a strong classifier. The strong classifiers are connected in series to form a cascade classifier, which is the ultimate score Cascaded classifiers can be used for face detection. Next, we will introduce AdaBoost algorithm and Haar features.

3.1. AdaBoost algorithm

The core idea of this algorithm is to train different weak classifiers with the same sample set to be tested. First, we set the maximum number of training cycles and initialize the sample weight, then train the weak classifier. When training weak classifiers, the error rate of weak classifiers is calculated first. Then select the appropriate threshold value to minimize the error; then update the weight, and send the new data to the next weak classifier for training. Finally, when the number of times is completed, several weak classifiers are obtained, which are added by the updated weight, and finally strong classifiers are obtained. Several strong classifiers are used to form a cascade classifier.

3.2. Haar characteristics

Haar features are mainly used to extract facial features. Haar features are divided into four categories which can be effectively combined into feature templates. There are just two kinds of rectangles that exist in the feature template: white and black. When extracting features, the template covers a certain area of the image, and then calculates the sum of the pixel values of two kinds of rectangles respectively. Finally, subtract the sum of the pixel values of two kinds of rectangles respectively to get the features of the template, which are Haar features. Since the facial features have bright information and there is a light dark relationship in local areas, it is very consistent with Haar characteristics.

On account of the reason that the feature template needs to calculate the sum of pixels for each region exist in image, which will result in a large number of repeated calculation of region pixel values. To solve this problem, we introduce the concept of integral graph. The integral graph is the sum of the pixels in the rectangular area formed from the starting point of the image to each point, as shown in the following formula

$$SAT(x,y) = \sum_{x_i \le x, y_i \le y}^{n} I(x_i, y_i) \tag{5}$$

Among them, SAT(x, y) is called the integral graph, and $I(x, y)$ is called the pixel value of the image (x, y) position. The integral diagram formula can also be calculated in an incremental way, as shown in formula (6)

$$SAT(x,y) = SAT(x,y-1) + SAT(x-1,y) + SAT(x-1,y-1) + I(x,y) \tag{6}$$

Through the way of integral graph, we can quickly get the features of a certain area. When face detection is carried out, Haar features can be used just like weak classifiers, and then combined with AdaBoost algorithm, multiple weak classifiers are combined into a strong classifier, finally the realization of the classifier is simple and the accuracy is high, and the detection effect is good.

4. Facial expression database

4.1. Acquisition of facial expression data set

In this paper, we use web crawler to get the original image data of facial expression. Web crawler is an automatic program for users to request website and extract data. Its basic principle is to grab the whole

page according to the specified URL link, then analyze, extract data information, and finally save the data information. Through crawling the network image data, 35887 facial expression pictures were collected, including 28709 pictures which come from the training set, 3589 pictures which come from the public test set and 3589 pictures which come from the private test set. Each image is composed of 400*500 gray-scale images with fixed size pixels. There are seven expressions: angry, disgusted, scared, happy, sad, surprised and neutral, We map them to digital labels 0-6 respectively. Figure 1 shows seven emoticons.

Figure 1. 7 facial expressions

4.2. Dataset expansion

When training the deep neural network, the generalization ability of the small sample set training model is relatively insufficient, and the persuasion is not enough when evaluating the network performance, so the artificial extended training data is considered. In this paper, the training set is expanded ten times by using the methods of turning transformation and translation transformation. Part of the operation renderings are shown in Figure 2.

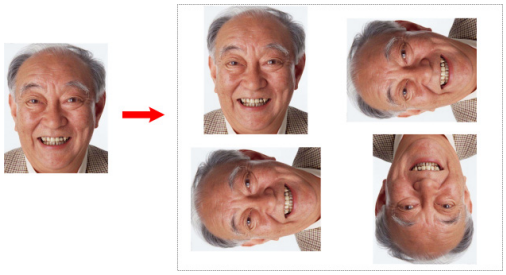


Figure 2. Partial operation effect

5. Experiment

In order to verify the performance of convolutional neural network model in face emotion recognition, this paper constructs a convolutional neural network model, and then randomly takes 80% of the data set as the training set model, and takes the remaining 20% of the data set as the test set performance. Figure 3 shows the recognition efficiency of the convolutional neural network model for the data set crawled by the network crawler.

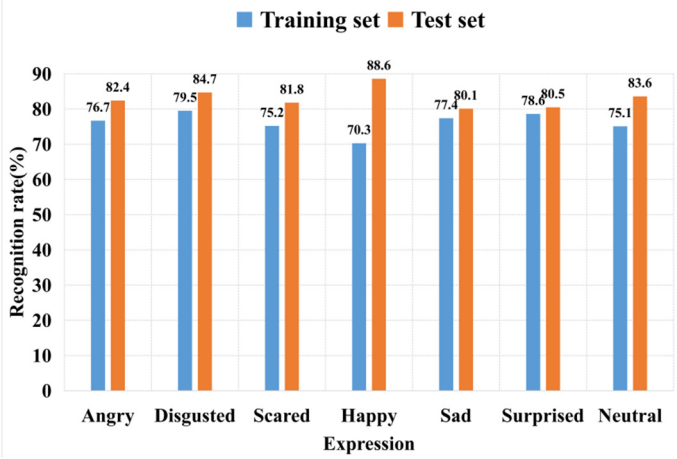


Figure 3. Experimental results

The experiment shows that the recognition rate of facial expression proposed in this paper is higher than 70% when applied to the training set and higher than 80% when applied to the test set, which has better performance.

6. Conclusion

This paper mainly studies the face expression recognition method based on convolution neural network. By constructing convolution neural network model, the face expression can be recognized and classified effectively. The application of facial expression recognition can call different model files for emotion recognition according to different target expression sets, which provides a theoretical and practical reference for facial expression recognition research. Considering the complexity of the system, this paper does not study the conditions of make-up and occlusion, and how to recognize facial expression under these extreme conditions needs further study. In addition, for convolutional neural network, it is necessary to collect as much data as possible and expand the data set reasonably, so that the trained network has better generalization performance and can reduces over fitting.

References

[1] Samson C, Blanc-Féraud L, Aubert G, et al. 2000 A Level Set Model for Image Classification. *International Journal of Computer Vision.* **40** 187-197.
[2] Liu Y, Guo J, Lee J. 2011 Halftone Image Classification Using LMS Algorithm and Naive Bayes. *IEEE Trans Image Process.* **20** 2837-2847.
[3] Moeskops P., Viergever M A, Mendrik, A M., et al. 2016 Automatic Segmentation of MR Brain Images With a Convolutional Neural Network. *IEEE Transactions on Medical Imaging.* **35** 1252-1261.
[4] Zhang L, Liu J, Zhang B, et al. 2019 Deep Cascade Model-Based Face Recognition: When Deep-Layered Learning Meets Small Data. *IEEE Transactions on Image Processing*. **29** 1016–1029.
[5] Lu J, Liong V, Wang, G, et al. 2015 Joint Feature Learning for Face Recognition. *IEEE Transactions on Information Forensics and Security*. **10** 1317–1383.
[6] Gao S, Zhang, Y, Jia K, et al. 2015 Single Sample Face Recognition via Learning Deep Supervised Autoencoders. *IEEE Transactions on Information Forensics and Security*. **10** 2108–2118.
[7] Lu J, Wang G, Zhou, J. 2017 Simultaneous Feature and Dictionary Learning for Image Set Based Face Recognition. *IEEE Transactions on Image Processing*. **26** 4042–4054.