Journal of Computer Science and Electrical Engineering	ISSN: 2663-1938	Volume 1 Issue 1 pp.1-7	March 2022

Research and Implementation of Face Expression Recognition and Classification Based on CNN

Qiu Haijing, Li Dan*, Zhang Kewen, Shi Yu, Chen Wen, Fan Shukang

Xuzhou University of Technology, Xuzhou, Jiangsu, China
Corresponding Authors' Email:3376486524@qq.com

Abstract

Facial expression is an important indicator to reflect human external performance and internal emotions and their changes. Studies have shown that facial expressions are the most commonly used and the most efficient method among the three ways of human dissemination of information: action, dialogue and expression. With the progress of the times and the development of society, artificial intelligence has gradually changed from theoretical scientific research to practical application into human daily life. The facial expression classification system in this paper uses the Python programming language combined with the Pycharm integrated development tool to develop the system, uses OpenceCV to preprocess the image, uses the Pytorch deep learning framework to build and train various neural network models, and combines Visdom to achieve data visualization.

Keywords: face recognition； expression recognition;cnn;

I. INTRODDUCION

Facial expression is an important indicator of human external performance and internal emotions and their changes. Research shows that facial expression is the most common and efficient method among the three ways of human communication: action, dialogue and expression[1]. Ekman, a famous psychologist, studied and analyzed the expressions of people of different races and cultures, and defined six basic expressions. As a result, the academic community generally started the research and exploration of machine recognition of facial expressions through the classification of these six basic expressions.

With the progress of the times and the development of society, AI has gradually changed from theoretical research to practical application into human daily life. At present, face recognition and expression recognition technology, as an important research direction of computer vision, have been widely used in the following fields:(1) In the field of distance education, the monitoring equipment reflects the students' classroom concentration and satisfaction with the course by identifying the students' facial expressions and body movements. Teachers can make adjustments to the subsequent courses and teaching through these data information, so as to better achieve the teaching purpose and improve the teaching quality and efficiency. (2) In the field of safe driving, the driver's facial expression is detected to determine whether there is fatigue driving or other dangerous situations, and timely warning is given to such dangerous situations to avoid traffic accidents and ensure the personal safety of passengers and drivers. (3) In the field of human-computer interaction, with the popularity of smart phones and other devices, machines can interact with humans more easily, and can also react to facial expressions more conveniently, such as adjusting the brightness of the device through the user's eye conditions, and pushing corresponding music in real time through judging emotions. (4) In the field of intelligent medical treatment, the patient's facial expression is detected to determine whether the patient will have an emergency or other unexpected conditions, and the medical staff on duty are warned in time to avoid the occurrence of unexpected conditions; Or make a better treatment plan according to the patient's mood changes at ordinary times.

With the rapid improvement of CPU and GPU computing power and the emergence of special chips in the past decade, artificial intelligence has developed rapidly. The method of applying deep learning methods to facial expression recognition has also been updated and iterated rapidly. Figure 1 shows the development of facial expression recognition algorithms since 2007.

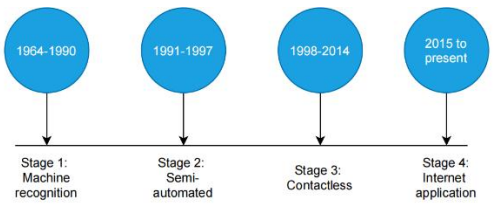


Fig.1. Development of facial expression recognition

II. INTRODUCTION TO CONVOLUTIONAL NEURAL NETWORK ALGORITHM

Convolutional neural network In the 1960s, Hubel et al. found this unique neural network structure by studying animal visual cells, which can effectively reduce the time required for neural feedback. In 1980, Fukushima et al. proposed the concept and implementation model of convolutional neural network based on this research. The

overall block diagram of YOLOv5 target detection algorithm is shown in the figure below.

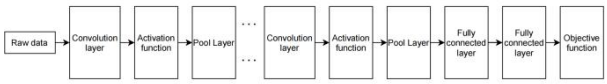


Fig.2.Convolution neural network infrastructure

A. Convolution neural network has three important mechanisms

(1) Local connection. Local connection can greatly reduce the use of parameters in the network. When processing high-dimensional inputs, it is difficult to make all neurons in each layer connect with each other, so only part of the area is connected with it. The connected space is called the receptive field of neurons.
(2) Weight sharing. Weight sharing is used to control the number of parameters in the convolution layer. Each filter is locally connected to the upper layer, and each filter that is locally connected uses a parameter. The advantage of this operation is that it can greatly reduce network parameters.
(3) Down sampling. Downsampling, that is, pooling, can maintain effective information in reducing feature dimensions. The function is to reduce the space size of data and speed up the operation.

B. Main composition of convolutional neural network

Convolutional neural network is mainly composed of convolution layer, pooling layer, full connection layer, Pooling layer can effectively reduce the size of the matrix, thereby reducing the parameters in the final full chain layer. Using pooling layer can not only speed up the calculation, but also prevent over fitting problems. There is also a filter in the pooling layer. However, the filter does not process the input data as convolution checks the input data for node weighted sum, but simply calculates the maximum or average value. The size of the filter, whether to fill all zeros, and the step size are also manually specified, but the depth is different from the convolution core depth. The filter used by the convolution layer spans the entire depth, while the filter used by the pooling layer only affects a node in depth. In the calculation process, the pooling layer filter must move not only in the length and width dimensions, but also in the depth dimension. The pooling layer using the maximum value operation is called the maximum pooling layer, which is used most frequently[3,4]. The pooling layer using the average value operation is called the average pooling layer, which is used less.
1. Maximize pooling. Select the maximum value output within the specified area, and the formula is:

$$y_{i,j}^{k} = \max_{0\le m,n\le s}\{x_{i*s+m,j*s+n}^{k}\} \quad (1)$$

2. Average pooling. Calculate the average value within the specified area for output. The specific calculation formula is:

$$y_{i,j}^{k} = \mathrm{mean}_{0\le m,n\le s}\{x_{i*s+m,j*s+n}^{k}\} \quad (2)$$

The full connection layer converts the feature map into category output. There are more than one full connection layer. To softening layer (softmax layer) and manipulation to achieve different functions. A convolutional neural network model with high efficiency and accuracy can be finally formed by reasonably setting the above layer structure and performing Dropout, NB and other operations between different layers as required.
The convolution layer is the most important layer in the entire neural network. The core part of this layer is the filter, or convolution kernel. The convolution kernel has two attributes: size and depth. The size commonly used is 3x3, 5x5, and also 11x11 convolution cores. The depth is generally understood as the number of convolution cores[2]. The size and depth of the convolution kernel are manually specified, while the weight parameters are randomly generated by the program during initialization, and these weight values are continuously optimized in the later training process to achieve the best classification effect. The process of convolution is to constantly multiply the RGB values of these images with these weight values to extract image data information. The specific calculation is shown in Figure 3.

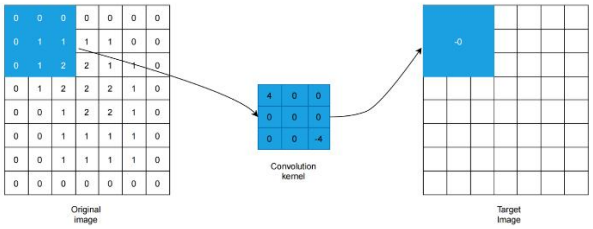


Fig.3. Calculation process of convolution layer

prevent over fitting, DropOut operation will be introduced between all full connection layers.
The Softmax layer does not belong to the unique structure level in CNN, and the results of image classification are output in the form of probability. All the outputs of the Softmax layer add up to 1, and the final category of the image is determined according to this probability.

C. Activation function

The activation function brings nonlinear changes to neural networks. We do not want to see a huge difference in results due to small changes. In order to avoid this situation, it is necessary to add nonlinear changes to the model. The nonlinear activation function makes up for the lack of expression ability of the linear model. The following describes several activation functions:
1. Sigmaid function

$$y = \frac{1}{1+e^{-x}} \quad (3)$$

2. Tanh function

$$y = \frac{e^{2x-1}}{e^{2x+1}} \quad (4)$$

3. ReLU function

$$y = \begin{cases} x & x > 0 \\ 0 & otherwise \end{cases} \quad (5)$$

Through practice, we know that the advantages and disadvantages of the three activation functions are as follows:

The output mapping of the sigmoid function is between (0,1), monotone and continuous, the output range is limited, and the optimization is stable. However, the power operation is expensive and prone to gradient dispersion; The Tanh function solves the problem of sigmoid's zero centered output. The derivative range becomes larger, between (0, 1) and sigmoid between (0, 0.25), and the gradient vanishing problem is alleviated, but the power operation, high calculation cost, and gradient vanishing problem; The ReLU function has unsaturated gradient and fast convergence speed. Compared with the sigmoid/tanh activation function, it greatly improves the problem of gradient disappearance and does not require exponential operation. Therefore, the operation speed is fast and the complexity is low, but it is very sensitive to parameter initialization and learning rate. There is neural element death, and the output mean value of ReLU is also greater than 0. Migration and neuron death will jointly affect the convergence of the network.

Compared with Sigmaid and tah, ReLU discards complex calculations (here it refers to power operation) and improves the operation speed; For deep networks, the problem of gradient disappearance is easy to occur in the process of back propagation of sigmoid and tanh functions; ReLU will make the output of some neurons zero, which will result in the sparsity of the network, reduce the interdependence of parameters, and ease the occurrence of over fitting problems.

III. EXPERIMENTAL PROCESS

A. Model building and training environment

We use Python language and its programming tool pycharm. A deep neural network Tf environment is built. PyCharm has the functions of general IDEs, such as debugging, syntax highlighting, project management, code jump, intelligent prompt, automatic completion, unit testing, and version control. Using python to write python makes the code implementation more intuitive and simple. The Python environment is deployed using the software miniconda to manage the Tf environment. The full name of the Tf environment is ensorFlow, a symbolic mathematical system based on data flow programming, which is widely used in the programming implementation of various machine learning algorithms. Tensorflow has a multi-level structure, which can be deployed in various servers, PC terminals and web pages, and supports GPU and TPU high-performance numerical computing. It is widely used in Google's internal product development and scientific research in various fields.

B. Selection and division of data sets

Through extensive collection and search, we can now obtain CK+, MMI and FER-2013 data sets for facial expression recognition. The three data sets all contain six basic facial expression states, including smile, cry， surprise and etc. The CK+and MMI data sets have a small amount of image data. Although some images in FER-2013 data set are indeed not accurate, their image specifications are small and can better adapt to the current software and hardware conditions. Therefore, FER-2013 is finally used as the research data set.

The FER-2013 dataset contains 35887 single channel grayscale images of facial expressions, including seven expressions: angry, disgusted, scared, happy, sad, surprised and neutral[5]. After that, the application of the model will process the collected image with gray level and then input it into the convolutional neural network model for analysis and judgment, which can solve the problem that facial expression recognition is vulnerable to light and color.

C. Preliminary design and configuration of model

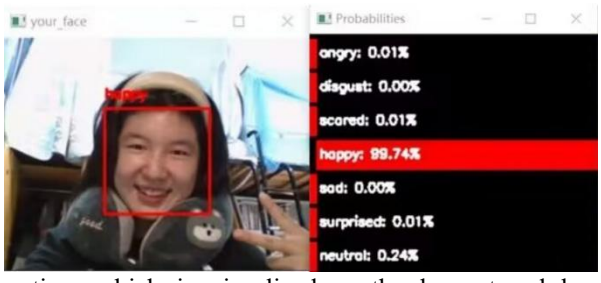
Preliminary design: The whole convolution neural network consists of three convolution segments, three full connection layers and a softening layer. Each convolution segment contains two convolution layers with the same convolution operation. The model is trained in batch_ size； The Adam optimizer is used to automatically adjust the learning rate during training; 3 * 3 convolution kernel is adopted for all convolution layers; Sigmoid activation function is used.

D. Practical application

After taking pictures from the camera, you can obtain pictures from the picture library, and the acquired pictures can be displayed in the software interface for identification. Perform image light compensation, image graying, Gaussian smoothing, histogram equalization, image contrast enhancement, binary transformation and other operations on the image[6]. The processed face image is located and the eyes, nose and mouth are marked for feature extraction. The feature values of eyes, nose and mouth are extracted from the located face image. The feature value extracted from the picture is compared with the value in the background database to complete the recognition function.

E. Expression recognition results

Figures 4 to 10 show the final running results of the model. Each person in the figure can clearly see what kind of emotion is on his face. The model also gives the highest probability value of his corresponding



emotion, which is visualized as the longest red bar. Different from other facial expressions, the probability of the seven expressions is visualized, which can more clearly reflect the user's emotions[7].

Fig.4.happy

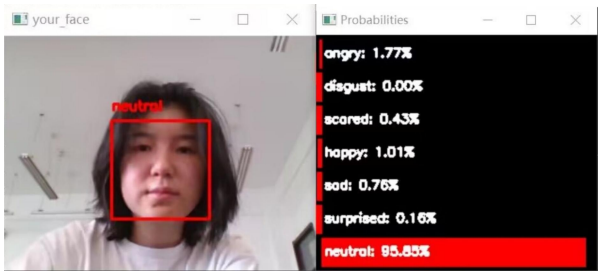


Fig.5. neutral

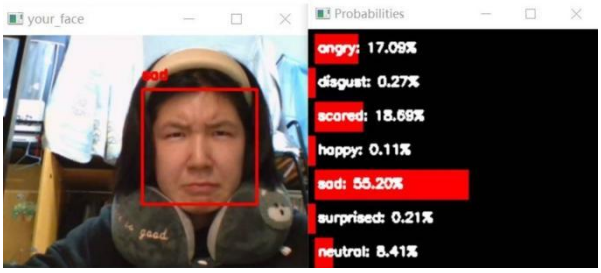


Fig.6. sad

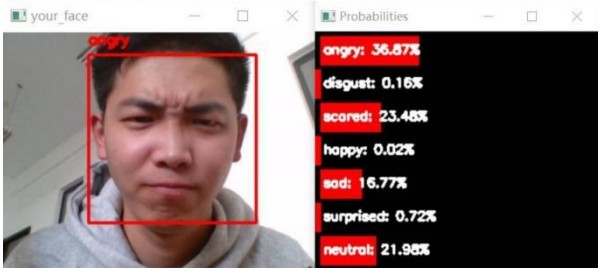


Fig.7. angry

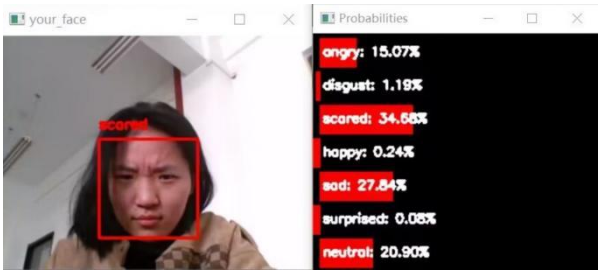


Fig.8. scared

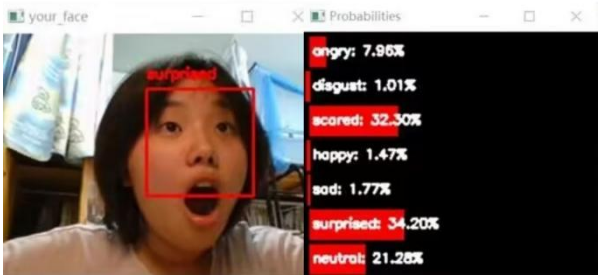


Fig.9. surprised

Fig.10.disgust

Figures 11 and 12 show facial expression recognition under different light intensities. There are some differences in the accuracy of the same expression under different light intensities.The accuracy of the expression recognition will be improved when the light intensity is good, and the accuracy will be decreased when the light intensity is weak, although the expression can still be recognized.

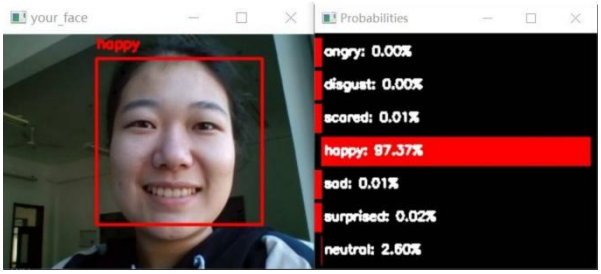


Fig.11. When the light intensity is good

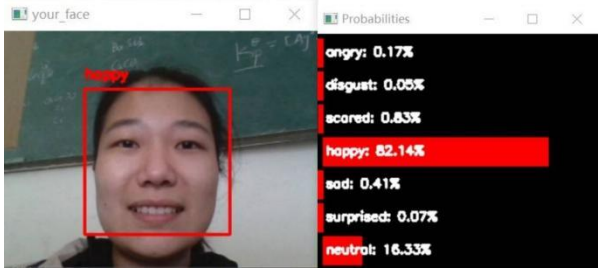


Fig.12.When the light intensity is weak

Figures 13 and 14 show facial expression recognition in different scenes [10]. In the indoor test, the accuracy will decline, which is not accurate enough, but in the outdoor test, the accuracy will be more accurate.

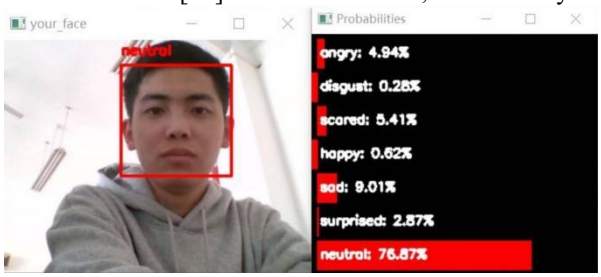


Fig.13. Indoor test results

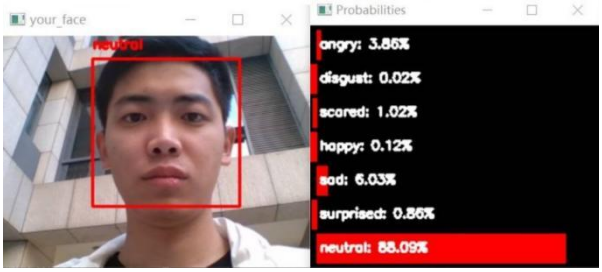


Fig.14.Outdoor test results

Figures 15 and 16 show facial expression recognition of different genders. The gender of the user was not distinguished when the model was trained with the test set, but the gap between male and female ratio was not large, so the influence of gender on the accuracy of the final test was not particularly obvious. The accuracy of male and female test subjects was almost the same when they had the same expression.

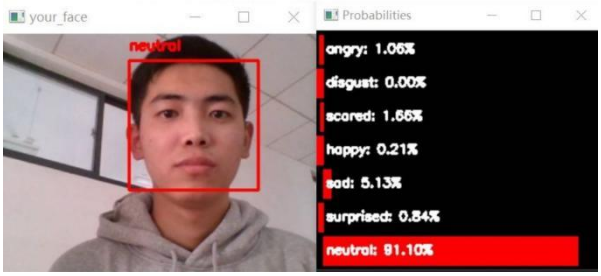


Fig.15. Male test results

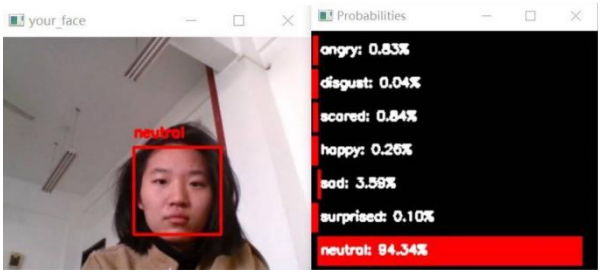


Fig.16. Female test results

Figures 15 and 16 show facial expression recognition with or without occlusion. During the test, it was found that whether or not to wear glasses had a great impact on the results, because the facial masks such as glasses and masks would hinder the extraction of facial feature values, and some key points could not be detected[11]. In addition, although there were no pictures of people wearing glasses in the training set, the number of pictures used was not large. These factors together led to a deviation in the test results

.

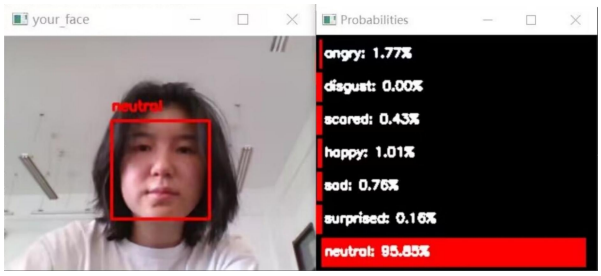


Fig.17. No obstructions on the face

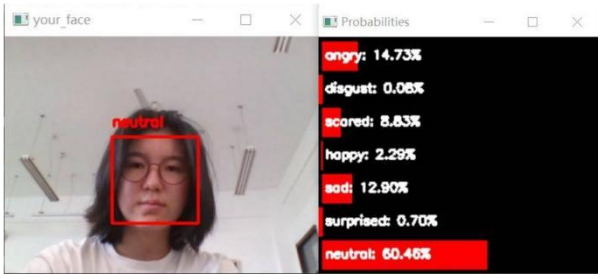


Fig.18. Face is covered

IV. Summary

This system is developed based on the Torch deep learning framework, learning and using convolutional neural network algorithm, and using a variety of framework technologies, including computer vision module, image processing module, data processing and analysis module, which makes the development process convenient. This paper mainly realizes the following work:

(1) We studied the traditional methods of face feature detection , facial expression classification and convolutional neural network methods, thought about and explored the differences between different processing methods and research routes, as well as the technical difficulties and bottlenecks faced by various technologies. After fully explaining and understanding the advantages and disadvantages of the general convolutional neural network, we conducted a comparative study of its data, making the research results more intuitive and pointing out the direction for future development.

(2) For the task of facial expression analysis, a facial expression recognition system including data preprocessing module, training and verification module and expression recognition module is implemented, and the optimal results are obtained when the system is completed. At the same time, in the process of applying theory to practice, we also learned the cutting-edge technology of contemporary computer vision，the use of various frameworks and functional modules. Moreover, through learning and practice, we have a general understanding of the theoretical knowledge and implementation process of these technologies.

(3) In the experimental practice, through detecting and recognizing the face and locating the key feature points of the face, we have completed the recognition of seven kinds of human expressions, which is the basic requirement of this system[12]. In addition, the effects of light , environment and facial occlusion on facial expression recognition are particularly studied. Increase more data in the data set that can be recognized by the system to improve the accuracy of expression recognition.

Facial expression recognition system can help better achieve human-computer interaction, fundamentally change the relationship between people and computers, so that computers can better serve human beings. In a word, facial expression recognition has great potential application value in the fields of psychology， intelligent robots , intelligent monitoring , virtual reality and synthetic animation. In this system, it helps to achieve better teaching communication between teachers and

students in online classes, and meets the current teaching requirements.

ACKNOWLEDGMENT

This work was supported in part by Jiangsu Provincial College Student Innovation and Entrepreneurship Training Program(Grant:xcx2022182)，Xuzhou Science and Technology Plan Project (Grant:KC21303), Jiangsu Industry University Research Cooperation Project (Grant:BY2021159), Jiangsu Educational Science "14th five year plan" Project(Grant:C-c/2021/01/65), the Sixth "333 project" of Jiangsu Province, Natural Science Research Projects of Colleges and Universities in Jiangsu Province(Grant: 22KJA520012).

REFERENCES

[1] Luo Xiangyun, Zhou Xiaohui, and Fu Kebo. "Face Expression Recognition Based onDeep Learning." Industrial Control Computer 30.5(2017):2.

[2] Yandong, L. I. , Z. Hao , and H. Lei . "Survey of convolutional neural network." Journal of Computer Applications (2016).

[3] Nguyen, H. , et al. "Facial expression recognition using a multi-level convolutional neural network." (2018).

[4] Yandong, L. I. , Z. Hao , and H. Lei . "Survey of convolutional neural network." Journal of Computer Applications (2016).

[5] Yi, J. , et al. "Facial Expression Recognition Based on t-SNE and AdaboostM2. " International Journal of Fuzzy Logic & Intelligent Systems 13.4(2013):315-323.

[6] Mcduff, D. , et al. "AFFDEX SDK:A Cross-Platform Real-Time MultiFace Expression Recognition Toolkit." ACM (2016).

[7] Hu, B. , Y. Huang , and B. Chen . "A novel facial expression recognition method based on semantic knowledge of analytical hierarchy process." Journal of Image and Graphics 16.3(2011):420-426.

[8] Liang-Hua, H. E. , et al. "The research advance of facial expression recognition." Journal of Circuits and Systems 10.1(2005):70-75.

[9] Jiang, B. , and K. Jia . "A Local Discriminative Component Analysis Algorithm for Facial Expression Recognition." Acta Electronica Sinica 42.1(2014):155-159.

[10] Liu, X. M. , H. C. Tan , and Y. J. Zhang . "New Research Advances in Facial Expression Recognition." Journal of Image and Graphics (2006).

[11] Ran, Wei , et al. "Facial Expression Recognition System Based on Multiple Feature Integration." Journal of Image and Graphics 14.5(2009):792-800.

[12] Zhengguang, X. U. , H. Yan , and L. Zhang . "Study of Independent Component Analysis Based on Facial Expression Recognition."Computer Engineering 32.24(2006):183-185.