Design and Implementation of Face Emotion Recognition System Based on CNN Mini_Xception Frameworks

Lili Sun¹, Chenhao Ge¹ and Yuanchang Zhong*¹

¹ School of Microelectronics and Communication Engineering, Chongqing University, Chongqing, 400000, China

*Corresponding author’s e-mail: zyc@cqu.edu.cn

Abstract. In the new era, the rapid development of information technology makes the face recognition technology in artificial intelligence also develop rapidly, among which, face expression recognition has become a research hotspot. In recent years, due to deep learning, convolutional neural network and multilayer perceptron and other related algorithms have become the research focus of scholars, so their wide application in the field of facial expression recognition is also the direction of face emotion recognition exploration and research. In addition, computer software is widely used in human daily life, so it is also very important to design and implement an intelligent, real-time and universally applicable UI interface for facial emotion recognition system. Therefore, this paper firstly completed the facial expression recognition model based on the Mini_Xception framework of CNN by training the FER2013 expression database. Secondly, the system UI interface is designed and realized through PyQT5, OpenCV, Keras and other libraries. The final results show that the system is based on the algorithm model, and the UI interface designed and realized can not only recognize the saved pictures, but also recognize the real-time face emotions through the camera, and the overall effect of the system is outstanding.

1. Introduction

In this era of rapid development of science and technology, artificial intelligence has become one of the research hotspots[1]. As a result, face expression recognition in artificial intelligence has become the research direction of many scholars. In everyday communication between people, there are usually two means: verbal and non-verbal[2]. Studies have shown that about 45 percent of human communication is verbal, while the remaining 55 percent is non-verbal, which is mainly expressed through facial emotions. Therefore, facial expression is a very necessary and important means of non-verbal communication in interpersonal communication, which is mainly used to show people's complex and changeable inner thoughts. Emotional expressions can be divided into seven categories: neutral, angry, disgusted, fearful, happy, sad and surprised. In many applications, these facial emotions can be used to convey signals that reflect human consciousness and mental states[3]. Therefore, it is necessary to design and implement a facial emotion recognition system to create a broad application space in the fields of psychological education, human-computer interaction, automatic driving and public safety.

Facial expression recognition is a very challenging topic in the field of biometric recognition and affective computing[4]. It was developed in recent decades, and now it has gradually become a research hotspot. Many institutions at home and abroad are doing research in this field. As far as facial expression recognition is concerned, it can be divided into two aspects: extraction and selection of facial expression features and classification methods. Feature extraction methods include principal component analysis...
(PCA) based on feature face, independent component analysis (ICA), wavelet transform (WT) and other methods to extract facial features. The classification methods mainly include support vector machine (SVM), hidden markov model (HMM), neural network (NN) and the deformation of neural networks[5].

In the past decade, an important breakthrough in the field of artificial intelligence is Deep Learning, which not only makes great progress in the application of images, sounds and texts, but also becomes a direction of the development of neural networks. Deep learning is a kind of machine learning, and machine learning is the only way to achieve artificial intelligence[6]. The concept of deep learning originates from the study of artificial neural network. The multi-layer perceptron with multiple hidden layers is a deep learning structure. Deep learning combines low-level features to form a more abstract high-level representation of attribute categories or features to discover distributed features of data. Back in the 1940s, neural networks sought to solve many of the problems of machine learning by mimicking brain cognition[7].

In addition, neural network is a machine learning architecture, in which all individual units are connected together in the way of weight, and weight parameters are obtained through network training, so it can be called neural network algorithm. The idea behind artificial neural network algorithms is to mimic the way the human brain thinks. The human brain receives input signals through the nervous system and then responds accordingly[8]. The researchers hope to simulate the brain's thinking in an abstract way. The neural network can learn the distribution law of big data and complete the whole learning process by using the back propagation algorithm. The back propagation algorithm calculates the parameter errors of each layer. The errors between layers interact with each other, and the network model is constantly revised and updated through the errors. Among them, CNN are specially designed according to the characteristics of data structure. CNN is the most biological network structure so far, and has excellent performance in many application fields[5].

Therefore, in recent years, the application of deep learning and CNN to images and videos has become the research direction of many scholars. At present, domestic research on the application of deep learning and CNN to face emotion recognition is just beginning[5]. The research on how to efficiently apply the methods of deep learning and CNN to facial expression recognition is not only an exploration of the research field of facial expression recognition, but also an exploration of the application field of deep learning and CNN. In addition, with the development of computer software and user interface design (UI), the design of facial emotion recognition into a visual interface or software has become a development direction[1].

To summarize, this paper uses Mini_Xception framework in CNN as the recognition model, and designs and implements a UI interface of face expression recognition system through PyQT5, OpenCV, Keras and other libraries. The system has high expression recognition accuracy and can track face expressions. It can not only recognize the saved pictures, but also recognize the real time facial emotions through the camera.

2. Design of facial expression recognition system

2.1. CNN and Mini_Xception Frame

It is well known that when fully connected network is used, the light intensity of each pixel in the input image is regarded as the value of the corresponding neuron in the input layer. For example, for a 28*28 pixel image, the number of input neurons is 784, leading to the high dimension of network parameters, which is difficult to train. In addition, the fully connected network architecture does not consider the spatial structure of images, and treats input pixels that are far away or close to each other on the exact same basis, making it difficult for the network to learn spatial structure information. The CNN adopts an architecture suitable for image characteristics, which makes the network model more consistent with data structure characteristics and is conducive to the training of deep and multi-layer networks. CNN is very good at image classification. The advantages of CNN are local connection, weight sharing and subsampling in space or time. These characteristics make the CNN have translation, scaling and distortion invariance to a certain extent[1].
Among them, Mini_Xception is the mainstream frame of CNN (As shown in Figure 1)[1]. Mini_Xception architecture is inspired by Xception architecture, which is another improvement to Inception V3 proposed by Google after Inception. Depthwise Convolution is also marked as Depthwise Convolution (S) to replace the convolution operations in Inception V3. Xception's network structure is slightly better than Inception V3 on the Imagenet dataset and significantly better than Inception V3 on image classification datasets containing 350 million images or more, while both structures maintain the same number of parameters, and the performance gain comes from more efficient use of model parameters.

The Mini_Xception architecture combines the use of residual modules and depth-separable convolution. The residual module modifies the expected mapping between the subsequent two layers, and the learned feature becomes the difference between the original and the expected feature. The Mini_Xception structure further reduces the number of parameters by eliminating the parameters in the convolutional layer on the basis of removing the final full connection layer. The structure has about 60,000 parameters, a reduction of 80 times compared to traditional CNN.

The Mini_Xception architecture is a full CNN consisting of four residual depth separable convolutions, each of which is followed by the addition of batch normalization and ReLU excitation functions. The final layer uses global average pooling and soft-max excitation functions to make predictions. The architecture achieved 95% accuracy in gender classification tasks. The accuracy of the sentiment classification task was 66% when tested on the FER-2013 dataset[3]. The final schema can be stored in an 855KB file. Therefore, this paper chooses this framework as the expression recognition model.

2.2. Step of facial expression recognition

The main steps of facial expression recognition (As shown in Figure 2) include pretreatment, feature extraction and selection, classification and subsequent processing. Since the data set FER2013 is more complete and more in line with the real life scene, this paper mainly selects the FER2013 training and test model. Preprocessing is to prevent the network from overfitting too fast, and some image transformations can be made artificially, such as flipping, rotating, cutting. The above operation is called data enhancement, and another major benefit of data operation is to expand the amount of data in the database, which makes the network robustness of training stronger. Feature extraction is to convert the pixel data of the shape, movement, color, texture and spatial structure of the face and its components into structured data that can carry out expression classification. It usually reduces the dimension of the input space, and the reduction of this dimension can reduce the burden of dimension. CNN's Mini_Xception architecture is used here for feature selection and classification recognition.
2.3. Overall system design

The whole system consists of neural network training and facial expression recognition. Neural network training is to study the parameters of the neural network based on neural network structure and shall be stored in a corresponding model file, the process requires a considerable amount of calculation, the hardware requirements for training equipment is very high, usually CPU training already could not reach the requirements of data training on hardware, so you need to use GPU to training.

The recognition of facial expressions is the core of the system. When the trained model file is obtained, in addition to using test set data to test the network training results, the more important thing is to use the corresponding code to complete the call to the model, so as to test whether the model can achieve the expected good results in the expression picture data under the real environment. The implementation of the whole system in software will be completed around the network training and expression recognition, using PyQT5, OpenCV, Keras and other libraries to complete the design of the application layer interface[1]. The overall framework is shown in Figure 3.

3. System UI interface design and implementation

3.1. System UI interface framework and composition

The research of facial expression recognition has made new development in the era of artificial intelligence, and the related achievements in this field have been greatly transformed into the industrial field, and better technical application direction has also been obtained. In view of the application requirements, it is necessary to build a human-computer interactive visual facial expression recognition system. Facial expression recognition system should have the following characteristics: easy to operate and visualization, high degree of intelligence, strong universality, strong real-time. The design process of the software system is shown in Figure 4.

This article UI interface, mainly includes picture entry, face positioning and face expression recognition and other functions[1]. The development environment of the interface is based on Windows 10 platform. Among them, this paper uses PyQT5, OpenCV, Keras and other libraries to design and implement a face emotion recognition system UI interface. The interface of the system is shown in Figure 5. The interface includes three button controls: select the model of emotion recognition, open the camera, and select the existing image file of expression. In addition, it also includes the recognition time, recognition results, the percentage of various emotions and so on.
Start
Read in the picture/Call the camera
Whether a face is detected
Yes
Facial expression recognition
No result return
End
Returns the identification result and the identification time
According to the results

Facial expression recognition

Figure 4 System software design flow chart.

Face Emotion recognition v1.0
v1 2021.07
Selection model
Time: 0 s
The live camera is on
Result: None
Choose picture

Figure 5 The UI interface diagram of the system.

Figure 6 Select the model function diagram.

Figure 7 Call the camera for real-time face emotion recognition function map.

Figure 8 Select the existing picture for emotion recognition function map.
3.2. **Display of the main functions of the system**

The system in this paper has three main functions, which are described in detail as follows.

Function 1: As shown in Figure 6, model file can be selected by clicking the icon indicated by the arrow, and then recognition can be made based on the model. In addition, which model is selected can be known from the contents in the green box. In this paper, the default model is selected.

Function 2: As shown in Figure 7, the camera can be turned on for real-time face emotion recognition by clicking the camera logo in front of the blue box, and the blue box can show whether the camera is on or not. The effect is shown in Figure 7.

Function 3: As shown in Figure 8, you can select an emotional picture of a face by clicking the folder mark in front of the black box, and then recognize the facial expression through the selected model. At this time, the yellow box in the picture shows that the camera is off. Therefore, this recognition method needs to be changed in the first and second places in the picture.

In a word, this system from the choice of network model, or from the design of UI interface are very good highlight its obvious advantages.

4. **Conclusion**

The purpose of this system is to make use of the Mini_Xception network based face expression recognition method studied in this paper to quickly locate the input face images or real-time faces through the UI interface of the system, and to recognize face expressions in time, and to display the recognition time, result and recognition proportion of each emotion. The comprehensive effect of the system is good. In the future, the recognition accuracy of the algorithm can be improved on this basis, and the improved algorithm model can be selected through the interface of the designed system for recognition, and the system can be applied to related fields.

**Acknowledgments**

This research was financially supported by the Fundamental Research Funds for the Central Universities (No. 2020CDCGTX055, No.2019CDCGTX302, No.2018CDPTCG000141). I would like to express my sincere thanks to the financial support of the project, to Professor Yuanchang Zhong for his improvement of this article, and to Junmei Yang and Chenhao Ge for his help.

**References**

[1] Michael R I, Sam E. A Survey on Human Face Expression Recognition Techniques[J]. Journal of King Saud University Computer & Information Sciences, 2018:S1319157818303379-.

[2] Ullman, S. (2019). Using neuroscience to develop artificial intelligence. Science 363(6428):692–693.

[3] Tian, Y. L. , Kanade, T., Cohn, J. F., Li, S. Z., & Jain, A. K.. (2005). Facial expression analysis. Handbook of Face Recognition. Springer London. 247–275.

[4] Li, T. H. S., Kuo, P. H., Tsai, T. N., & Luan, P. C. . (2019). Cnn and lstm based facial expression analysis model for a humanoid robot. IEEE Access, PP(99), 1-1.

[5] Siqueira, H., Magg, S., & Wermer, S. . (2020). Efficient facial feature learning with wide ensemble-based convolutional neural networks. arXiv.

[6] Nguyen, B. T., Trinh, M. H., Phan, T. V., & Nguyen, H. D.. (2017). An efficient real-time emotion detection using camera and facial landmarks. Seventh International Conference on Information Science & Technology. IEEE. 2017.7926765.

[7] Meshgi, K., Oba, S., & Ishii, S.. (2018). Efficient Diverse Ensemble for Discriminative Co-Tracking. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE.

[8] Zhang, K., Huang, Y., Du, Y., & Wang, L.. (2017). Facial expression recognition based on deep evolutional spatial-temporal networks. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 26(9):4193-4203.