International Journal of Applied Science and Engineering Review

IJASER

ISSN: 2582-6271

Vol.2 No.4; July-Aug 2021

RESEARCH ON EMOTION RECOGNITION BASED ON FACIAL EXPRESSIONS

Yukuan He and Hongli Zhu

School of Information and Electronic Engineering, Zhejiang University City College, Hangzhou, China

DOI: http://dx.doi.org/10.52267/IJASER.2021.2404

ABSTRACT

With the development of science and technology, emotion recognition has been integrated into every aspect of our life, whether in medical services, public safety or in human-computer interaction, the demand for emotion recognition technology is expanding, therefore, in recent years, emotion recognition based on facial expressions is still a research hotspot. Traditional methods for emotion recognition by extracting extracted feature points and feature classification require a lot of manual operations, and the accuracy and precision are not high. In this paper, we adopt the algorithm of deep learning of convolutional neural network to train the images. the data set used is FER2013, the system can get the picture in real time through the camera for recognition, or by reading the input The experimental results show that the deep learning algorithm for emotion recognition research has higher accuracy and robustness.

KEYWORDS: facial expression, emotion recognition, convolutional neural network, deep learning, accuracy

1. INTRODUCTION

The most important module of the system designed in this paper is the face emotion recognition, and the model building of emotion recognition is a crucial work, because the good or bad recognition effect will be directly related to the quality of the system. The expression recognition model used in this paper is the mainstream framework mini_Xception model of CNN. Firstly, we introduce the knowledge of deep separable convolution, then we introduce the mini_Xception neural network framework, and then we train and tune the mini_Xception network to build the model of emotion recognition.

2. Deeply separable convolution

In this paper, we use the mainstream CNN framework, mini_Xception. Xception is another improvement to Inception v3 proposed by Google after Inception, and its core is the deep separable convolution operation. By using deep separable convolution, it can reduce the computation and network parameters and save CPU computing space than standard convolution.

International Journal of Applied Science and Engineering Review

IJASER

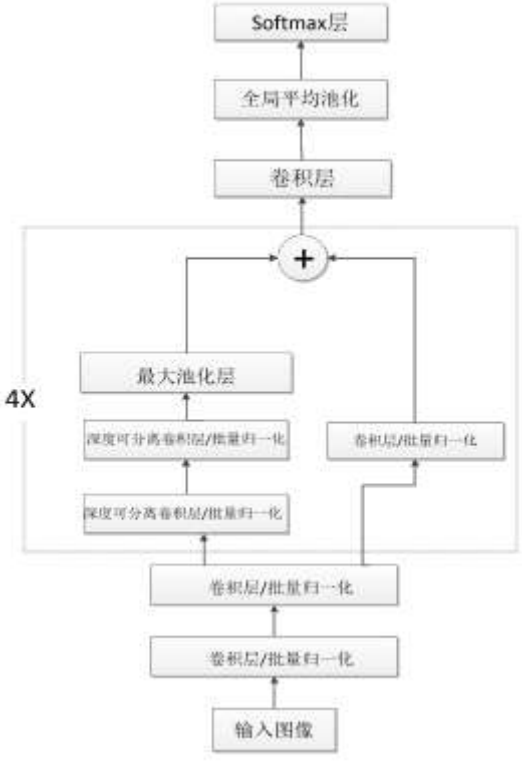
ISSN: 2582-6271

Vol.2 No.4; July-Aug 2021

2.1 Lightweight convolutional neural network mini_Xception

The mini_Xception network architecture is a model proposed by Arriaga et al. in 2015 for face expression classification, which has the advantages of high accuracy and good real-time performance. mini_Xception network introduces the ideas of deep separable convolution and deep residual network, and the introduction of deep separable convolution operation can greatly reduce the network parameters, while the introduction of residual module The introduction of the residual module speeds up the convergence speed of the network and enhances the learning ability of the network for subtle features such as facial expressions.

mini_Xception is a fully convolutional neural network, which is a modified version of the Xception network, using the name "mini" because of its relatively small number of network parameters. The architecture of the model is shown in Figure 1, which contains four residual depth-separable convolutional modules, that is, the ordinary convolutional operations in the residual module are replaced with depth-separable convolutional operations. The last convolutional layer is not followed by a fully connected layer, but by a global average pooling layer, which prevents the network from overfitting and further reduces the network parameters; after that, the softmax layer is connected to carry out the emotion classification work.



ISSN: 2582-6271

Vol.2 No.4; July-Aug 2021

Figure 1: mini_Xception network structure diagram

The model architecture reduces the model parameters to 60,000 by deeply separable operations, and the number of parameters is 391 times lower than that of the Xception network. Besides, the model architecture was tested on the FER2013 expression dataset and obtained an accuracy of about 66% in the emotion recognition task, and the weights of the model can be stored in 886KB files.

In this paper, we use the deep learning framework TensorFlow, in which the data representation is tensor, and the main data is "tensor flow", where "tensor" refers to the high-dimensional arrays in the process of operation, and the "flow" is the process of computation and transformation between high-dimensional arrays. Therefore, TensorFlow is an algorithmic framework for representing the deep learning process through computational graphs.

The variable name is the unique identifier of a tensor. In the deep network structure, the common variable name can be given by the method "node: srcoutput", where node is the name of the node and srcoutput indicates the first output of the current tensor from the node, while the tensor stream is just read into the When the tensor stream is just read into the network, there is no output at that time, so the srcoutput value is 0.

When defining the network structure of mini_Xception in Python code, the statement to read in the tensor stream is img_input = Input(input_shape), and the input variable name of the network should be "Input: 0" according to what was described earlier. When defining the structure of the network, the last layer needs to output the tensor stream, and the statement of the output result is: output = Activation ('softmax', name='predictions') (x), so that we can get the output variable name "predictions/ softmax ", which gives the variable names of the input and output of the mini_Xception network.

2.2 Building an expression recognition model based on mini_Xception

The previous section illustrated the basic features of the neural network architecture mini_Xception, and now we take fixed scenes as an example to illustrate the basic contents of expression recognition model building, including model tuning and participation testing, in preparation for the software implementation of the recognition system later. Four expression data sets were introduced above, and FER2013 is used as an example in this paper.

2.3 Model Tuning Participation Test

When tuning the mini_Xception neural network, in the setting of batch_size, the GPU can perform better for a batch of powers of 2. Therefore, it tends to perform better when it is set to 13, 32, 64, 125, etc. than

ISSN: 2582-6271

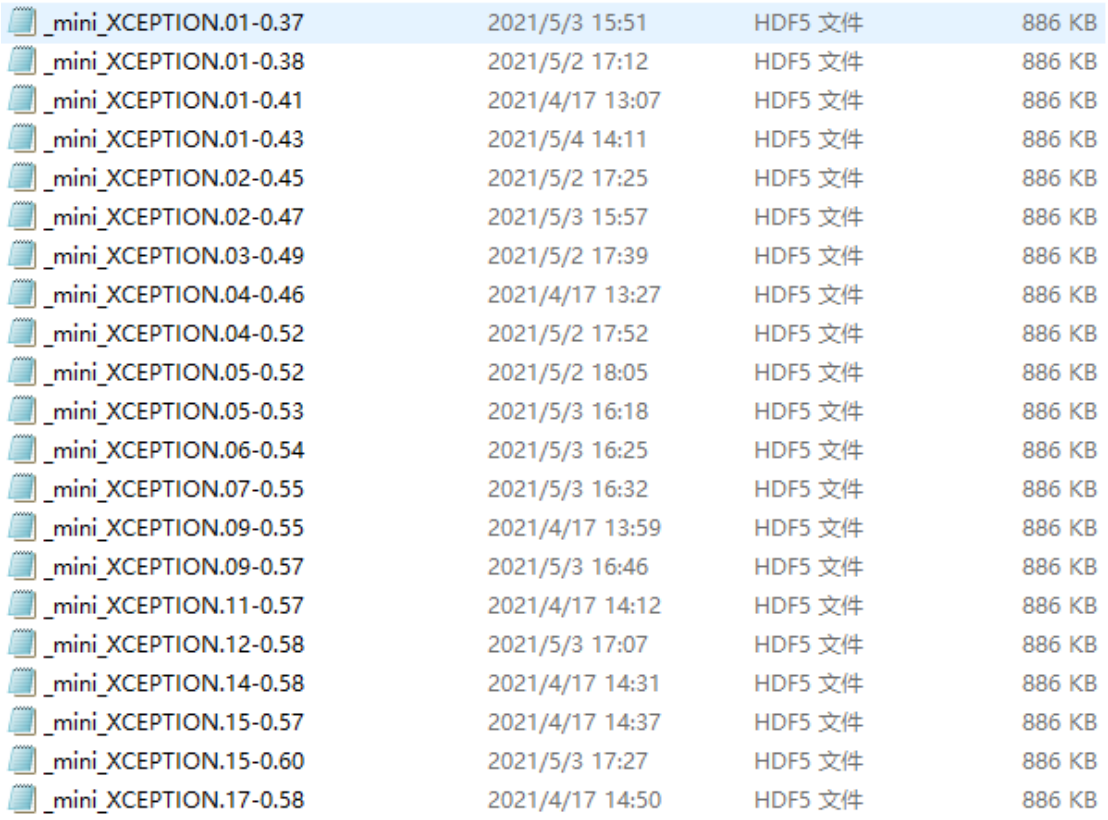
Vol.2 No.4; July-Aug 2021

when it is set to a multiple of whole 10 or whole 100, and the epoch measurement setting will also have an impact on the accuracy rate. The main parameter settings are shown in Table 1.

Table 1: Partial parameter setting table of network model

batch_size	*verbose*	*patience*	*num_epochs*	*validation_split*
32	1	10000	50	0.2

After setting the relevant parameters, the model can be trained. After each round of training, the accuracy and loss rate of the training set and the accuracy and loss rate of the validation set will be output, and the loss rate of the validation set after this round of training is lower than the loss rate of the validation set in the previous round, and the model file hdf5 file will be generated and saved to the specified path, as shown in Figure 2, the generated hdf5 model file can save the structure of the network, weights and other information. The naming rule of this file is: the first number indicates the number of training rounds, and the second number indicates the accuracy of the validation set after the end of this round of training. From the generated file, it can be concluded that the results can vary even when trained in the same number of rounds.



ISSN: 2582-6271

Vol.2 No.4; July-Aug 2021

Figure 2: Screenshot of the generated model file after training

3. Experimental Result

The emotion recognition module is the core module of this study. The four previous modules of picture-video acquisition module, video pre-processing module, face detection module, and expression recognition module are all paving the way for the emotion recognition module.

(1) Loading model

In this paper, we call the cv2.CascadeClassifier() method to load the face detection model and the load_model() method to load the face expression recognition model, and we should pay attention to the setting of relevant parameters when loading the model. In the above, it is mentioned that the detected images are normalized to a size of 48*48, so the code also needs to define the input image size as 48.

(2) Read in images or videos to obtain results

As mentioned earlier, the faces array temporarily stores the face images, and here the elements in the faces array need to be taken out in turn and fed into the expression classifier to classify each element with sorted() function.

The above is the design and implementation of each system of the software, the next is to test the system on the actual scenario, and finally, according to the test results, point out the advantages and shortcomings of the system. The interface of the whole system is shown in Figure 3. As can be seen from the figure, we can carry out real-time recognition by turning on the camera, and we can also load pictures and carry out recognition on pictures as shown in Figure 4. When performing face emotion recognition you can choose your own trained model, the predicted results of different models in the same scene will also vary, as shown in Figure 5, the predicted result with a model with 37% accuracy in the test set is scare, while the predicted result with a model with 65% accuracy is surprise, the actual expression in this figure should be surprise. interface The right half is the probability of predicting various emotions during recognition, and the one with the highest probability is the final recognition result.

International Journal of Applied Science and Engineering Review

IJASER

ISSN: 2582-6271

Vol.2 No.4; July-Aug 2021

Figure 3: System interface

Figure 4: Real-time camera recognition and loading image results

International Journal of Applied Science and Engineering Review

IJASER

ISSN: 2582-6271

Vol.2 No.4; July-Aug 2021

Figure 5: Differences in recognition results of different models

4. CONCLUSION

This paper develops and programs the software system with the help of toolkits such as OpenCV and Python3.6 programming language, and introduces each module in detail. When testing the system, for the system to directly read the graph surface for emotion recognition, it can more accurately identify the current emotion; but for the camera to identify emotion in real time, it is influenced by environmental factors, and the recognition results have a large difference with the actual, and cannot predict the current expression very accurately for the following reasons: first, influenced by the FER2013 data set, with mini _Xception neural network training out of the accuracy ceiling is not high; second, the use of the computer comes with the camera resolution is not high, cannot clearly obtain the real-time picture.

REFERENCE

[1] Ekman P, Friesen w V. Facial Action Coding System (FACS): a tec hnique for the measurement of facial actions [J]. Rivista Di Psichiatria,1978,47(2):126-38.

[2] Waller B M, Smith Pasqualini M. Analysing facial expression using the facial action coding sy stem (FACS[J]. Mouton De Gru yter,2013:917-931.

[3] Huanhuan Wang. Research on Emotion Recognition Based on Facial Expressions D]. Harbin Institute of Technology, 2017.

[4] Lv Zhen. Research on video-based facial expression recognition technology [D]. Qingdao University of Technology,2013.

[5] Wu Q, Shen Xunbing, Fu Xiaolan. Research on micro-expressions and their applications[J]. Advances in Psychological Science, 2010, l8(9):1359-1368.

[6] Li Xia, Lu Guanming, Yan Jingjie, et al. A review of multimodal dimensional sentiment prediction[J]. Journal of Automation, 2018, 44(12): 2142-2159.

[7] Cortes C, Vapnik V. Support-vector networks [J]. Machine learning, 1995,20(3): 273-297.