EXTRACTION OF EMOTIONAL FEATURES OF HUMAN FACE

Yu He and Hongli Zhu

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

DOI: <http://dx.doi.org/10.52267/IJASER.2021.2403>

ABSTRACT

This paper introduces the face recognition dataset and the feature extraction techniques. and the deep learning method for face recognition will be introduced. The specific Pycharm software development platform is used, the system program is implemented by Keras, Open Cv, PyQt5 and other libraries, and the mini_XCEPITIOM convolutional neural network developed by Google is used to classify face emotions, The system can recognize the face by getting the picture from the camera in real time or by reading the input face, and the accuracy rate is about 65%.

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

1. INTRODUCTION

In recognition problems, effective features can help us discover the essence of things, therefore, the extraction of expression features is an important element in face emotion recognition. The extraction methods of face emotion features mainly include two kinds: traditional methods and deep learning methods. Among them, traditional methods are relatively mature and have achieved good results in general face emotion recognition; deep learning methods have played a great role in recent years and can better overcome the challenges of emotion recognition in complex environments. Both methods have their advantages, and this paper will introduce the deep learning method for face recognition.

2. Face emotion recognition dataset

Currently, the existing publicly available face emotion recognition datasets are relatively few and of small order of magnitude. There are two main types of commonly used face emotion recognition datasets: one is the field dataset, which provides enough training samples for deep learning to be carried out; the other is the controlled dataset in a laboratory setting, which generally has a smaller specification. In this section, I will select four typical representatives from each of these two types of datasets to introduce them.

Meanwhile, for the convenience of expression, I will replace the various expression categories mentioned in the previous article with the corresponding abbreviations, namely AN: angry, HA: happiness, SA: sadness, SU: surprise, DI: disgust, FE: fear, and CO: contempt and NE: neutral.

2.1 JAFFE data set

Japanese Female Facial Expressions (JAFFE), a controlled dataset in a laboratory setting, contains 213 face images, consisting of seven types of expression samples from 10 Japanese women, and for each type of expression there are three to four images for each individual. The difficulty in recognition of this dataset is that the sample size is too small and the individual expression states can vary widely from subject to subject, as shown in Figure 2.1, where the first female has a small change in facial amplitude under different emotions, while the second female has a large change in facial amplitude under different emotions. In addition, there are a small number of labeling errors in this dataset, as shown in Figure 2.2.

Figure 1: Performance of different subjects with the same expression

Figure 2: Example with error labels

2.2 CK+ dataset

The extended Cohn-Kanada (CK+) dataset was collected by P. Lucy. The dataset contains 593 video sequences of 327 tagged expression picture sequences from 123 subjects, with subjects from an ethnically diverse group of adults. The video sequences show the transition of faces from neutral to peak expressions, and all sequences provide the full FACS encoding of peak frames. The expression picture sequences were divided into seven expressions: normal, angry, contempt, disgust, fear, happy, and sad. For each picture sequence, only the last frame was provided with an expression tag, so a total of 327 images were tagged, and the specific data composition is shown in Table 1.

Table 1: Number of images of each type of expression in the CK+ dataset

<i>Expression Category</i>	<i>AN</i>	<i>DI</i>	<i>FE</i>	<i>HA</i>	<i>SA</i>	<i>SU</i>	<i>CO</i>
Number of sequences	45	59	25	69	28	83	18

The CK+ dataset has a wide range of applications, both for the detection of facial motion units AU and for static and dynamic face emotion recognition. This author used AAM face tracking, extracted appearance features and geometric features, and combined with SVM to construct a baseline system, but did not provide a specified training and test set, so the evaluation algorithms on this dataset are not uniform. For static face emotion recognition, the common data selection method is to extract a number of frames containing peaks in each sequence. Also, many methods exclude the CO category from such performance evaluation because of its small sample size and high recognition difficulty.

3. Deep feature extraction based on convolutional neural network

Deep learning methods are more flexible compared to traditional methods. When using convolutional neural networks to process face images, the undersized and large number of convolutional kernels are equivalent to detection operators, and they are determined by training a large number of samples, so that deep feature extraction is the result of automatic learning.

Figure 3: Schematic diagram of convolutional neural network

3.1 Network unit

As shown in Figure 3, the basic units of a convolutional neural network are: a convolutional layer, a pooling layer, and a fully connected layer. The convolutional layer has a set of learnable convolutional kernels and is therefore able to perform convolutional operations on the input image to generate a specific activation feature map. The local connectivity in the convolutional operation guides the network to learn correlations between adjacent pixels, while the weight sharing mechanism of the convolutional kernels greatly reduces the number of parameters for training.

3.2 Activation function and loss function

The purpose of the activation function is to introduce nonlinearity into the neural network, and there are generally three types, as shown in Figure 4. Among them, the sigmoid function has the point of simple derivation and stable output, but its existence of the saturation region easily leads to training non-convergence; tanh function converges faster than the sigmoid function, but the calculation is more time-consuming; ReLU function can increase the washout to a certain extent, while making the gradient disappearance phenomenon effectively moderated, and is also the activation function used in this study.

Figure 4: Common activation functions and the corresponding graphs

In addition to the typical Hinge loss and mean square error loss, there are some new losses that have applications in the face expression recognition problem. For the classification problem in this paper, it is more effective to choose a multiclassification logarithmic loss function.

3.3 Optimizer

The common optimizers in deep learning are Adagrad, RMSProp, SGD, Adam, etc. The Adagrad algorithm uses an adaptive approach to adjust the learning rate, but there is a problem of learning rate disappearance, and it is mainly suitable for scenarios with sparse features. The RMSProp algorithm is optimized on the basis of Adagrad by introducing decay coefficients to prevent premature termination of training. SGD generally refers to the small batch gradient descent method, which has the advantages of both batch gradient descent and stochastic gradient descent, but the disadvantage is that the learning rate needs to be set manually and the learning rate is consistent for all parameters, which easily leads to the function falling into local limit values. The Adam algorithm, on the other hand, is a much better and improved version, and it has user-friendly features in addition to fast and stable convergence, because its hyperparameters basically do not need to be adjusted, so Adam's algorithm is used in this paper for all network parameter optimization.

4. Face detection module

Emotion recognition is built on the basis of face recognition, so this module also needs to use the OpenCV computer vision library. Face detection mainly includes: loading the cascade detection file, inputting the grayscale processed image and judging whether there is a face, then returning the detection result to the original color image, framing the face image in the color image and storing it temporarily. The flow chart of face detection is shown in Figure 5.

Figure 5: Flow chart of face detection

As mentioned above, the image is grayed out and stored in the gray variable, and then the detectMultiScale() function in OpenCV is called to implement face detection, which is the core method of face detection, and some important parameters need to be set.

Table 2 Introduction of the main parameters of the detectMultiScale() function

<i>Parameter Name</i>	<i>Note</i>
Image	The input image, in this paper, is the grayscale image gray stored temporarily earlier
objects	Detect the face target sequence and return the rectangular box coordinates
Scale Factor	The ratio of each image size change, set to 1.1 in this paper
Min Neighbors	The number of eligible objects in the domain is set to 5 in this paper.
flags	This article is set to cv2.CASCADE_SCALE_IMAGE
minimize	The minimum range for object detection, set to (30,30) in this paper

The detection result is the position of the face in the picture and the result without the face part respectively. If the face is detected, the rectangle box is drawn and the coordinates of the upper left and lower right corners of the rectangle box need to be calculated before drawing, so that the size of the rectangle box can be mastered for drawing, and then the rectangle () function is called to get the coordinates and the detection box can be drawn. After the above work is done, the face image can be stored, and the faces array is defined in this paper for temporary storage of face images.

REFERENCES

- [1] Cortes C, Vapnik V. Support-vector networks [J]. Machine learning, 1995,20(3): 273-297.
- [2] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]. International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates, 2012:1097-1105.
- [3] Oliver R F, Berni J, Viallet J E, et al. A Fast and Accurate Face Detector Based on Neural Networks[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001, 23 (1):42-53.
- [4] Ma Y, Ding X. Real-time multi-view face detection and pose estimation based on cost-sensitive AdaBoost[J]. Tsinghua Science and Technology, 2005,10(2):152-157.
- [5] Shi X, Shan S, Kan M, et al. Real-time rotation-invariant face detection with progressive calibration networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:2295-2303.
- [6] Viola P, Jones M J. Robust real-time face detection [J]. International journal of computer vision, 2004,57(2): 137-154.