World Journal of Information Technology
ISSN: 2959-9636
DOI: 10.61784/wjit231105

EXPRESSION RECOGNITION SYSTEM BASED ON CONVOLUTIONAL NEURAL NETWORK

Huihui Wang, Dan Li*, Ruiqun Xu, Hengjia Zhang, Yi Liu, Bohua Li
Xuzhou University of Technology, Xuzhou, Jiangsu, China.
Corresponding Authors' Email: 1347334157@qq.com

Abstract: With the continuous development of artificial intelligence technology, student expression recognition in the classroom has become an important research direction in the field of education. However, existing expression recognition methods often have problems such as low classification accuracy and high recognition difficulty, making it difficult to meet the needs of practical applications. In order to solve these problems, this paper proposes a method for student classroom expression recognition based on convolutional neural network. By collecting images of students' classroom expressions and using technologies such as preprocessing, feature extraction, and model training, we can accurately identify students' classroom expressions, monitor students' status in real time, and remind teachers to change the classroom atmosphere to help students adjust in time to improve learning. efficiency, while also further promoting the development and application of emotional education.
Keywords: Artificial intelligence; Deep convolutional neural network; Expression recognition; Classroom status recognition

1. INTRODDUCION

In the era of smart education, the application of technology in the field of education has made significant progress. The topic of this project focuses on the research of student classroom expression recognition technology based on deep learning in the context of smart education to improve the quality and effect of education. In recent years, the application of facial recognition related technologies in security, payment, attendance, etc. has made great progress, such as smart bus payment systems [1]. In the student classroom, we face a series of challenges, such as the expression of students' emotions, their participation, and identifying whether students' emotions are concentrated or dispersed. This requires high-precision expression recognition technology, and it is this challenge that stimulates our research interest.

Generally, traditional face recognition related technologies are mainly based on face recognition of visible light images in practical applications. Although this method has been studied for more than thirty years, this technical method still has some insurmountable shortcomings, especially when the lighting conditions are insufficient, or when the ambient lighting changes, the recognition effect of human facial features will drop significantly, and may even fail to meet the requirements of actual system recognition [2]. The student classroom expression recognition system in this article is implemented based on the convolutional neural network algorithm, which has long been used in various fields as a mainstream research method. The detection accuracy is higher than traditional methods, exceeding 95% [3]. Compared with traditional face recognition technology, face recognition technology based on convolutional neural networks has obvious non-contact application advantages, better concealment advantages, higher security advantages and high cost-effective advantages [4].

At present, student expression recognition needs to be achieved with the help of facial expression recognition method combining ResNet18 and capsule network, residual network facial expression recognition method, etc. Traditional methods are prone to misjudgment or missed identification when performing face recognition, which affects the recognition rate. However, this study uses a convolutional neural network algorithm, which has the advantages of good expansibility, stability, and scalability. It can adapt well to various complex network environments. Compared with general expression recognition systems, this system uses real-time image or video monitoring technology to automatically identify students' behavior and performance in class. It detects students' expressions to see if they are tired, distracted or blinking. By analyzing these data, the system can help teachers better understand students' learning status and performance, and promptly discover learning difficulties and problems students may encounter. This provides teachers with valuable information that allows them to provide more personalized educational services to each student. By discovering and solving students' problems in a timely manner, teachers can guide students' learning more effectively and improve their learning outcomes and satisfaction. The application potential of this system is very wide in the field of education, helping to create a positive learning environment and improving the quality and effectiveness of education.

The simplified process of the entire student classroom expression recognition system is shown in Figure 1.

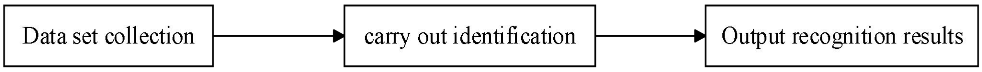

Figure 1. Simplified Flowchart of Classroom Expression Recognition System

2. Introduction to FaceNet algorithm

The method adopted by FaceNet [5] is to map images into Euclidean space by directly training a deep convolutional neural network , and image similarity is directly related to spatial distance. The FaceNet model designs a multi-layer convolutional network to map facial image features to a 128-dimensional vector space , thereby achieving the main function of mapping a facial image to a 128-dimensional vector space. FaceNet uses a deep convolutional neural network as the backbone network, and its accuracy on the LFW data set can reach 99.63%. The FaceNet algorithm deep learning algorithm is used to efficiently extract facial features and coordinate the ambiguity of biometric features with the accuracy of the cryptographic system [6].

The overall block diagram of the FaceNet algorithm is shown in Figure 2 below:

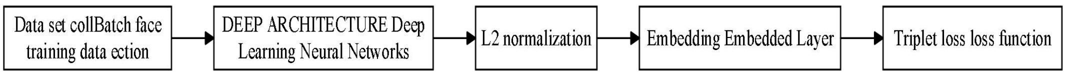

Figure 2. FaceNet network structure diagram

This model uses the deep learning neural network structure to put the face training data of a photo into it. After deep convolution, a set of face data will be obtained. This set of data will then be L2 normalized and embedded. After layer processing, the 128-dimensional face feature vector of this face can be obtained, and then training can begin.

This article uses ordinary convolution and depth-separable convolution for feature extraction. Compared with ordinary convolution, the depth-separable convolution structure block can reduce the parameters of the model. Therefore, this article uses the mobilenetV1 network structure as the backbone feature extraction network [7]. As shown in Figure 3.

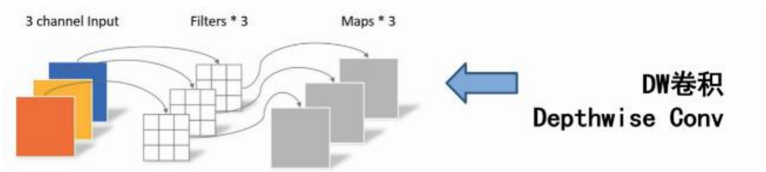

Figure 3. Schematic diagram of depth separable convolution feature extraction

There are three photos in a triplet triplet. Anchor and Positive represent photos of the same person, while Anchor and Negative represent photos of different people. Using the triplet loss (anchor, positive, negative) method to train this model, the Euclidean distance between (Anchor, Positive) becomes smaller, and the Euclidean distance between (Anchor, Negative) Getting a larger distance is the desired result of training. Therefore, a value needs to be set to adjust the Euclidean distance between the two, satisfying the following conditions:

$$\left\| f(x_i^a) - f(x_i^p) \right\|_2^2 + \alpha < \left\| f(x_i^a) - f(x_i^n) \right\|_2^2 \quad (1)$$

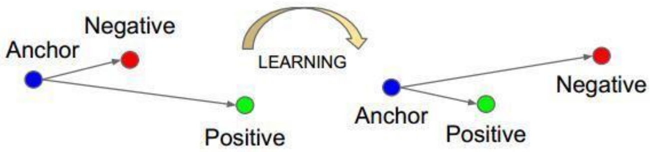

Figure 4. Triplet loss method training diagram

From Figure 4, in the formula, x_i^a represents the randomly selected data Anchor in the training set, x_i^p represents the sample of the same type as the Anchor selected in the training set, and x_i^n represents the sample selected in the training set that is not of the same type as the Anchor. By deforming the above formula, the loss function of the model can be obtained:

$$loss = \sum_{i=1}^{N} \left[\left\| f(x_i^a) - f(x_i^p) \right\|_2^2 - \left\| f(x_i^a) - f(x_i^n) \right\|_2^2 + \alpha \right] \quad (2)$$

3. experiment procedure

3.1 Data Set Collection

The design of the face recognition system based on convolutional neural network designed in this article is shown in Figure 5. First, images of the face to be recognized are collected. The collected images need to be preprocessed. The preprocessed images are sent to the convolutional neural network. CNN performs processing and finally performs face recognition.

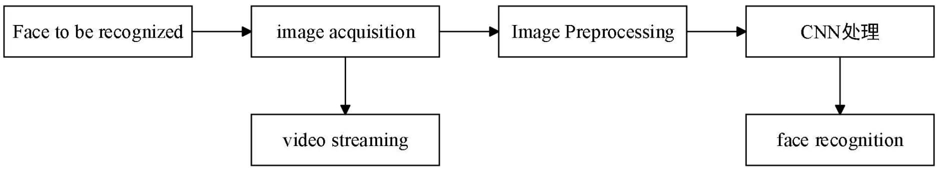

Figure 5. Face recognition process

3.2 Image Preprocessing

The data set used in this article is the CASIA - WebFace data set, and image preprocessing is now performed on it.
Image data preprocessing is to process the sample into a data type that can be read by the convolutional neural network through a series of operations such as adjusting the image size and scaling, dividing the data set, one-hot encoding label vectorization, and pixel normalization. Data preprocessing includes image resizing, cross-validation method partitioning the data set, and pixel normalization.

3.3.1 Adjust image size

The size and proportion of the image data are adjusted because two fully connected layers are added to the convolutional neural network designed in this program. In the network configuration, the input of the fully connected layer has a fixed dimension, but the core operation of the fully connected layer is Matrix-vector product, that is, multiplying the input feature map matrix by the corresponding weight matrix. The dot product operation between matrices requires the same two matrix dimensions to fix the input data feature map matrix. This means that if the size and proportion of the image data are not adjusted uniformly, the program will report an error and be unable to run.

3.3.2 Cross-validation method to divide the data set

Multiple models have been obtained through the above process, and the model finally selected must be the model with strong generalization ability and the best effect on unknown new data. Therefore, before training the model, the data mastered must be divided. Strictly speaking Generally, it is divided into training set, verification set and test set. The model is trained on the training set, the effect is tested on the verification set, parameter settings are adjusted, and the final test is performed on the test set. In order to ensure the final effect, these three sets cannot overlap. A common ratio is 8:1:1. Of course, it is usually possible to only have a training set and a test set. The sample data set used previously only has a few hundred, so there is no verification set divided. The train_test_spli used in this article belongs to the hold-out method, which randomly uses a part of the data as the training set and the rest as the test set.

3.3.3 Pixel normalization

The normalization of pixels is achieved by dividing the original pixel value by the maximum value of all original pixel values. The maximum pixel value is generally 255 and the minimum value is 0. The formula is as follows:

$$y = \frac{X - X_{-min}}{X_{-max} - X_{-min}} \quad (3)$$

X_min in Equation 1 is 0. y is the normalized value, X_max represents the maximum value of the sample, and X_min represents the minimum value of the sample. When the pixels of the normalized image are in the range of 0~1.0, they are still between 0~255, and the image is still valid.

Through the above steps, a self-built face image data set for model training is obtained. Some examples are as Figure 6:

Figure 6. Some image examples

3.3 Student Classroom Recognition System based on Convolutional Neural Network

The student classroom identification system based on convolutional neural network is a system that realizes automatic identification and analysis of students' behavior and performance in the classroom. The system can monitor classroom images or videos in real time, and automatically identify students' behaviors and performances in class, such as expressions, fatigue detection, looking left and right, etc., thereby helping teachers better understand students' learning status and performance, and discover students' problems in a timely manner. Learning difficulties and problems in order to provide more personalized educational services. The initial interface of the system is divided into three parts. The left side is mainly the camera's control panel, the middle is the system's detection results, and the right side is the image and the corresponding module of the detection results. The control panel includes the camera's open and close buttons, and the middle image The module displays the original image and the recognition result image successively. The UI interface is as shown in Figure 7:

Figure 7. Recognition results

Click to open the camera on the system interface: The user clicks a button on the system's graphical user interface (GUI) to open the camera function. After the camera is turned on, the user will face the camera and the system will capture the user's facial image through facial recognition technology. Click the Detect Face button. After capturing the facial image, the user can click a button to start the face detection function. The system will recognize the images input by the camera and match them with existing data sets. Once a matching facial image is found, the system will display relevant information to the user. Clicking the Detect Expression button will activate the expression recognition function. The system will analyze facial expressions, identify the user's emotional state, and display the analysis results to the user. For example, emotions such as "happy", "sad", and "angry " may be displayed. Label. Click on fatigue detection, and the system will analyze the user's eye status, count the number of blinks, and accumulate them to a preset threshold. The system will notify the user and pop up a " tired" warning to remind the user that they may need to rest. If the system detects that the user is shaking his head, it will trigger a "look left and right" warning, which may mean that the user is trying to avoid looking or doing other impermissible behaviors. The recognized images will be synchronously saved to the storage path of the training model: whether after face recognition, expression recognition or fatigue detection, the system will save the recognition results and corresponding images, and these data will be used for subsequent training. model to improve the recognition accuracy of the system. Click the close button and the system will turn off the camera and return to the initial interface.

3.3.4 Identification

During the classroom expression recognition process, the classroom expression recognition system based on FaceNet collects expression images from the camera, recognizes the facial expression images on the picture and marks the status. The rectangle is positioned as the face, and the blue font in the upper left corner indicates the expression state. The shape features are relatively stable, and there is no problem of being unable to capture the face due to insufficient light. Therefore, we use the feature that the shape of the logo is not easily affected by light and search for shape features. Detect whether a face is captured. If there is a face on the recognition image, the face will be framed by the anchor point , and the corresponding expression, fatigue index, whether to look left or right and other information of the face will be displayed above the anchor point frame to complete the recognition.

3.3.5 Output Recognition Results

Based on the model's input images and previous training data, specific recognition results for facial expressions, fatigue, and head orientation can be obtained. These results can be discrete categories, such as "happy," "sad," or "tired," or they can be continuous values representing levels of fatigue. Depending on specific needs, these results can be further used for decision-making or applications in other systems. In our experiments, we used 7 different facial expression markers as recognition objects, many of which had very similar images, such as laughing, grinning, and wry smiles. The marked patterns are all very similar, which makes identification very difficult. Our dataset also includes some less common keypoint recognition, such as fatigue detection, left and right gaze, and other types not included in other experiments. Due to search difficulties, the number of instances of these gesture types is small and the network cannot fully understand the characteristics of this expression. This dataset uses CK+(Cohn Kanade): a facial expression database containing images of various facial expressions and corresponding labels. Each sample has seven basic expressions (such as anger, sadness, disgust, etc.), as well as a neutral expression and driver fatigue dataset: This dataset contains video and eye feature data collected from driver behavior for fatigue detection. Based on video frames and eye movement data, it can be determined whether the driver is in a fatigue state and head direction dataset: This dataset is used for head direction or posture estimation and includes facial images at different angles and positions. Output the recognition results, as shown in Figure 8, Figure 9 and Figure 10.

Figure 8. Fatigue test results

Figure 9. Results for Shake Your Head Recognition

Figure 10. Results for Happy Recognition

4. Summary

Of recognizing student expressions in classroom based on convolutional neural network is continuously improving in recognition rate. At present, there are still some challenges in the research on face recognition methods based on convolutional neural networks. One of them is the problem of insufficient data. The lack of sufficient training data may cause the performance of the model to decrease. Additionally, overreliance on sequence data can also be a limiting factor, as access to continuous facial expression data can be limited. This requires building a complete expression recognition system and integrating advanced image preprocessing technology and expression detection technology to accurately extract facial expression features and make accurate recognition. In addition, for the problem of facial image degradation caused by environmental factors, image enhancement methods also need to be studied to solve problems such as insufficient light or uneven illumination. Face recognition methods based on convolutional neural networks are expected to be further improved and optimized. With more research and practice, we can expect the emergence of more efficient and accurate technical models for student expression recognition in the classroom. This will provide teachers with more powerful support to better understand students' learning status and emotions in the classroom, thereby promoting the development of personalized educational services.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

ACKNOWLEDGMENT

This work was supported in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant: 23KJA520013), Xuzhou Engineering College Student Innovation and Entrepreneurship Training Program Project (Grant: xcx2023193), the Sixth "333 project" of Jiangsu Province.

REFERENCES

[1] Jianhou G, Juxiang Z, Wenkai N, et al. An Optimization Algorithm for the Uncertainties of Classroom Expression Recognition Based on SCN[J]. International Journal of Software Science and Computational Intelligence (IJSSCI), 2022,14(1).

[2] Zhihui Z,M. JF, Lluis MG. Facial expression recognition in virtual reality environments: challenges and opportunities [J]. Frontiers in Psychology, 2023,14.

[3] Monica LL, Cenerini C, Vollero L, et al. Development of a Universal Validation Protocol and an Open-Source Database for Multi-Contextual Facial Expression Recognition[J]. Sensors, 2023, 23(20).

[4] Ivana K, Simon S, John MB, et al. Towards smart glasses for facial expression recognition using OMG and machine learning[J]. Scientific Reports, 2023,13(1).

[5] Qianyi Z, Baolin L. Construction of the brain-inspired computing model verified by spatiotemporal correspondence between the hierarchical computation of the model and the complex multi-stage processing of the human brain during facial expression recognition[J]. Applied Intelligence, 2023,53(21).

[6] Faten K ,Hela L. Neural style transfer generative adversarial network (NST-GAN) for facial expression recognition[J]. International Journal of Multimedia Information Retrieval, 2023,12(2).

[7] Sami RA, Hossein M M , Amirhassan M, et al. Dataset classification: An efficient feature extraction approach for grammatical facial expression recognition[J]. Computers and Electrical Engineering, 2023,110.