International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

Research on Learning Status Evaluation Based on Facial Expression Recognition

Shuqin Huang[1], Yong Xu[1], Xiaoyan Shi[1]
[1]School of Management Science and Engineering, Anhui University of Finance & Economics, Bengbu 233030, China
Email address: 393054429@qq.com

Abstract**—***Learning status directly affects the learning effect. Therefore, evaluating learning status is an important aspect of understanding students' learning status in teaching. Facial expressions are an important reflection of a person's inner world and can directly reflect the quality of learning status. This paper analyzes students' facial expressions based on a deep learning framework with attention mechanism, which has practical significance for assisting teachers in teaching and adjusting teaching plans and teaching content in a timely manner.*

Keywords**—***Facial expression recognition, learning status, attention mechanism, deep learning*.

I. INTRODUCTION

Facial expression is one of the most direct signals of a person's inner emotions and is a reflection of his or her inner world. The goal of Facial Expression Recognition (FER) is to identify the emotional state of an individual by analyzing facial images or videos[1]. The learning state directly affects the learning effect. Assessing the learning state is an important aspect of understanding students' learning status in teaching. Therefore, analyzing students' facial expressions is an effective way to understand students' inner world and learning state.

Facial expression recognition is essentially a classification problem. Machine learning, such as support vector machines(SVMs) and random forests, have shown excellent performance in classification. However, facial expression is a complex problem that is affected by various factors such as lighting, angle, and clarity, so facial expression analysis is very challenging. Deep learning has been widely used in facial expression classification due to its powerful automatic feature extraction and computational efficiency.

Most of these worked perform reasonably well on datasets of images captured in a controlled condition. Facial expression emotion analysis based on data under controlled conditions often cannot truly reflect students' true emotions. This paper uses deep learning to perform emotion analysis based on actual classroom video data, and then masters students' learning status, which is of great significance to improving learning efficiency and teaching effectiveness.

II. RELATED WORK

Facial expression recognition technology is more and more favored by researchers, and gradually developed. Because of the complexity of the recognition environment and the diversity of facial expression, and human expression is also related to psychology, compared with other biological recognition technologies such as face recognition and fingerprint recognition, facial expression recognition technology is not widely used and adoped. However, because of technology has a very important research value in human-computer interaction, and facial expression is the most direct and effective emotion recognition, facial expression become a hot research. Many research are committed to the research of expression recognition, and have achieved some research results.

As for the facial expression feature extraction methods, there are the histogram of oriented gradients (HOG)[2], SIFT[3], local binary patterns (LBP)[4], etc. HOG describes the features of an image by calculating the output of a series of gradients at the points of the image. The SIFT algorithm detects key points using the Difference of Gaussian (DoG) method, assigns an orientation to each key point, and generates descriptors based on the local gradients around the key points. LBP captures the spatial structure of an image by comparing the value of each pixel with its neighboring pixels.

In terms of research algorithm， many facial expression are studied by deep learning. Among many deep learning models, Convolutional Neural Network (CNN) is the most popular network model. In CNN-based approaches, the input image is convolved through a filter collection in the convolution lay. Khanzada et al. [5] took a deep dive, implementing multiple deep learning models for facial expression recognition. They not only improved the accuracy but also applied it to real world. Mohamad Nezami et al.[6] presented a deep learning model to improve engagement recognition from images that overcomes the data sparsity challenge by pre-training on readily available basic facial expression data, before training on specialized engagement data.

Some scholars integrated attention to analyze the facial expression. Minaee et al.[7] proposed a deep learning approach based on attentional convolutional network, which was able to focus on important parts of the face, and achieved significant improvement over previous models on multiple datasets. They also used a visualization technique which was able to find important face regions for detecting different emotions, based on the classifier's output. Li et al.[8] proposed a novel LBAN-IL for FER. LBAN include two operations which were local binary standard layer and encoder-decoder module. Local binary standard layer derived from local binary convolution prevented excessive sparseness of feature maps and reduced the number of learnable parameters. Encoder-decoder module generated attention-aware features and accurately discovered local changes in the face. There are other scholars using other methods to research facial expression. Barman et al.[9] calculated from normalized distance and shape signature pair to supplement the enhanced feature set fed into a MLP to arrive at different expression categories.

In terms of research data form, facial expression recognition has the recognition of static image and dynamic sequence according to whether it uses frame or video image[10]. Static FER relies solely on static facial features obtained by extracting handcrafted features from selected peak expression frames of image sequence. Dynamic FER utilizes spatio-temporal features to capture the expression dynamics in facial expression sequences. People research facial expression via image, audio, video and sensor. Kahou et al.[11] applied CNNs for extracting visual features accompanied by audio features in a multi-modal data representation. Some scholars research the 3D data source to analyze the facial expression. Chang et al.[12] described an ExpNet CNN for estimating 3D facial expression coefficients.

The emotion state and cognitive level affect the student's learning effect in education big data. Face Expression can well reflect the emotional state and cognitive level of students. Most current psychoanalytic methods tend to focus on attention and ignore the role of emotion in human learning. In view of this, Xu et al.[13]presented a multi-task learning implementation with a cascaded convolutional neural network (CNN) for face detection. Lee et al.[14] presented a focus-based assessment (PFA) approach using the concept of deep neural networks and a sequence of facial images. Although scholars have applied facial expression recognition to the field of education, there are still many problems worth studying.

III. METHODOLOGY

A. Brief Introduction of the Algorithm

We collected class videos, split them into images by frame, and filtered the images. In order to ensure the effectiveness of facial expression analysis, blurred images must be clarified. In addition, we removed non face images from the Fer2013 dataset. First, face detection is performed based on haar-like feature. Then, we have established a mini_xception_attention model based on mini_xception[15]. Extract facial features, and determine facial expressions and emotions with this model. Finally, facial expression and emotion analysis is applied to the evaluation of learning status. The haar-like feature determines whether the region may contain a face by calculating the grayscale difference of pixels in different regions of the image. At the same time, haar-like also uses a cascade classifier method to divide an image into multiple regions, and each region is gradually detected and screened to improve detection efficiency and accuracy. The mini_xception_attention model uses deep separable convolution to reduce computational complexity, controls the weight of channels with attention mechanism, while retaining gradient information through residual connections. Multiple module stacking enables the model to extract more complex features

B. Framework of mini_xception_attention

The mini_xception_attention model consists of four layers: input layer, basic convolution layer, module layer and output layer.

The basic convolution layer contains 2 ConV2D/BN/ReLU/Attention. Each module in the module layer is composed of two parts. One part is composed of the main path(mainP), including Sep_CONV2D/BN/ReLU /Attention/Sep_ConV2D/BN/MaxPool2D, and the other part is composed of the residual term (rsd), including ConV2D/BN, that is, module=concatenation(mainP,rsd). The module layer consists of four modules, each with different convolution kernels, and finally an output layer, which contains ConV2D/GlobalAvgPooling2D/Softmax. The model structure is shown in Fig. 1.

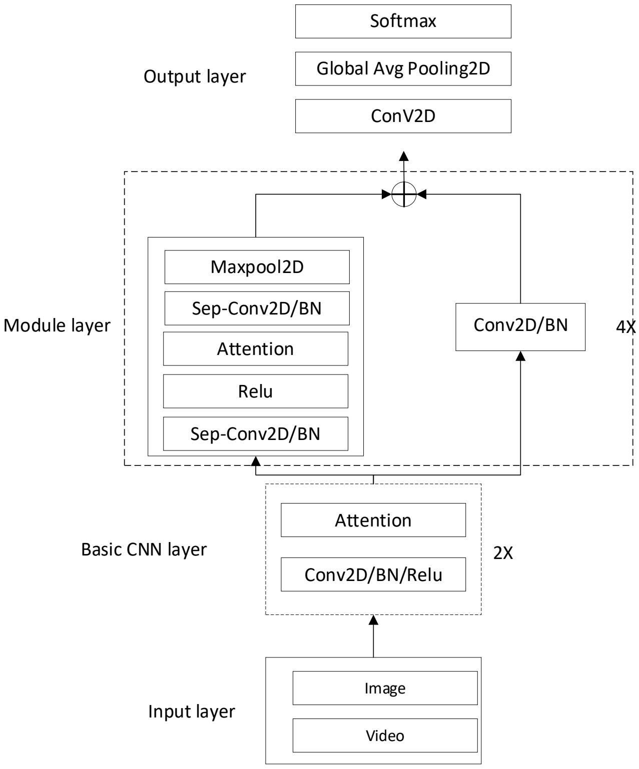

Fig. 1. Mini_Xception model with attention

C. Feature Extraction

⑴Haar Feature

Haar feature[16] is calculated through integral graph. The Haar feature extraction method is as follows.

Assuming that the four vertices of region D are a, b, c, and d, the sum of the pixels in region D is:

$$\boldsymbol{D_{sum} = SAT(a) + SAT(d) - SAT(b) - SAT(\ c)}$$

Let the coordinates of point d be (x, y), then

$$\mathbf{SAT(x,y)} = \sum_{x' \leq x, y' \leq y} \boldsymbol{I(x',y')}$$

Among them, I(x,y) represents the pixel value at the (x,y) position of the image, and the pixel point in the integral image stores the sum of all the pixel values to its upper left.

⑵Feature Extraction of mini_xception_attention

Image feature extraction is key. First, use the convolution operation to extract features. Let i be the input image, K be the convolution kernel, k be the convolution kernel size, c be the input channel, C be the number of channels, f be the output channel, then the output feature X is,

$$\boldsymbol{X(i,j,f)=\sum_{c=0}^{C-1}\sum_{m=0}^{k-1}\sum_{n=0}^{k-1}I(i+m,j+n,c)\cdot K(m,n,c,f)}$$

Then it goes through BatchNormalization, ReLU and attention. Suppose the output of the ReLU is r(i,j,f) , then the output after attention is $\mathrm{a(i,j,f)}=\mathrm{r(i,j,f)}\cdot\alpha_f$, where α_f is the channel weight calculated by the attention mechanism. Repeats this operation twice, and then further extracts features through the module layer to obtain feature Y.

⑶*Learning status evaluation system framework*

The process of the learning status evaluation system is basically to split the video image into static images, then perform facial clarity processing on the obtained images, establish a deep leaning model for the obtained clear images and train the model to obtain the final sentiment analysis results. The model is applied to the judgment and analysis of the learning status of individual students and the overall learning status of students. The overall learning status system process is shown in Fig. 2.

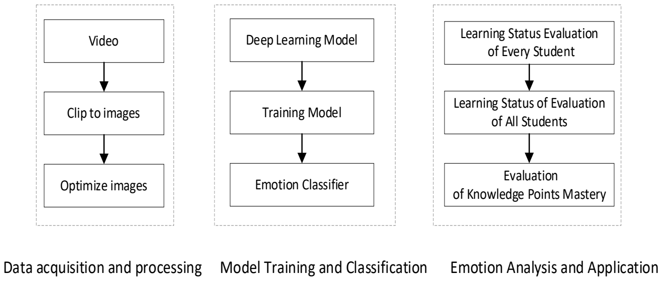
Video
Clip to images
Optimize images
Deep Learning Model
Training Model
Emotion Classifier
Learning Status Evaluation of Every Student
Learning Status of Evaluation of All Students
Evaluation of Knowledge Points Mastery
Data acquisition and processing
Model Training and Classification
Emotion Analysis and Application

Fig. 2. Framework of Learning Status Evaluation System

IV. EXPERIMENT

A. Category Definition of Learning State

In order to better represent the learning status, we classified the seven types of emotions. When a student is judged to be angry, fear, disgust, or sad, we consider this to be a negative emotion and that his/her learning status is not good. If a student is judged as having other emotional polarities, we consider this to be a positive emotion and that his/her learning status is good. We set a threshold to count the emotional polarity of a student over a period of time. If the number of students judged as 'angry ', 'fear', 'disgust', or 'sad' is greater than the threshold, it is considered that the student's learning status during this period is not good. Otherwise, it is considered that the student's status is good. If the emotional polarity of all students in the class is statistically analyzed over a period of time, the overall learning status of all students can be determined.

B. Evaluation of individual student's learning status

We divide the video into images by frame, and filter images to remove non student facial images and blurry images. We analyze the facial expressions of specific students, judge the facial expressions of students one by one, count the proportion of positive and negative emotional analysis, and determine the learning effect of students.

In addition, we can divide video by knowledge points. For example, for a 3-minute knowledge point, we analyze the emotional signals of students in the corresponding images of the video to understand each student's mastery of this knowledge point, so as to better adjust the teaching content and teaching methods. Some examples of emotional analysis of individual students are shown in Table 1.

TABLE 1. Samples of Single Student Learning Status Analysis

Students No.	Proportion of negative state	Proportion of positive state	Learning state
1	4.75%	95.25%	Good
2	59.17%	40.83%	Not Good

C. Overall evaluation of learning status

When we focus on a specific class, we collect videos of students in class, split into images, judge the facial expressions of all students in the images, count the percentages of positive and negative polarity, and then average the percentages of positive and negative polarity of all images to determine the overall learning status of all students. Perform multiple statistics according to different set standards, and then process them by averaging. The set standards can be divided according to the entire class or by knowledge points. For example, for the entire class, if the students' learning situation is counted by time, if the positive proportion is high, for example, greater than 50%,it means that the class is more effective. Otherwise, the teacher should find the reasons in a targeted manner and adjust the teaching plan.

You can also make statistics by knowledge point, so that you can have a more detailed understanding of students' mastery of the knowledge points. For example, a knowledge point is taught for 6 minutes, we can split the video within these 6 minutes, analyze and count students' emotions to understand their knowledge mastery status. We can set a threshold. When the positive signal is greater than this threshold, we think that the student's mastery is OK. Otherwise, we think that is not OK. The teacher needs to adjust the teaching progress or teaching methods. The sample of knowledge point mastery evaluation is shown in Table 2.

TABLE 2. Samples of Evaluation of Knowledge Points Mastery

Knowledge points	Proportion of negative signal	Proportion of positive signal	Whether mastered
1	36.41%	63.59%	OK
2	54.07%	45.93%	Not OK

V. CONCLUSION

Using deep learning models to recognize students' emotions can assist teachers in assessing students' learning status. By analyzing students' emotional states according to their needs, teachers can gain a better understanding of their students, enabling timely adjustments to teaching methods and content. This approach is crucial for improving teaching quality and enhancing students' learning efficiency.

ACKNOWLEDGEMENTS

This work was supported in part by the Natural Science Foundation of the Higher Education Institutions of Anhui Province under Grant KJ2021A0484.

References

[1] J. Yu, Z.H. Wei, Z.P. Cai, G.P. Zhao, Z.R. Zhang, Y.Q. Wang, G.C. Xie , "Exploring Facial Expression Recognition through Semi-Supervised Pre-training and Temporal Modeling, " *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* ,pp.4880-4887*, 2024*.

[2] V. A. Saeed, " A Framework for Recognition of Facial Expression Using HOG Features. International Journal of Mathematics," Statistics, and Computer Science, Vol. 2,pp.1-8, 2024.

[3] N. Ali, K.B. Bajwa,R. Sablatnig, S. A. Chatzichristofis, , Z. Iqbal, M. Rashid and H. A. Habib, "A novel image retrieval based on visual words integration of SIFT and SURF," PLoS ONE, vol.11,No.6:e0157428,2016.

[4] C.F. Shan, S.G. Gong and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," Image and vision Computing, vol.27,No.6,pp. 803-816, 2009.

[5] A. Khanzada, C. Bai and F.T. Celepcikay, "Facial Expression Recognition with Deep Learning," arXiv preprint arXiv:2004.11823, 2020.

[6] O. Mohamad Nezami, M.Dras, L. Hamey, D. Richards , S. Wan, and C. Paris, "Automatic recognition of student engagement using deep learning and facial expression," *Joint european conference on machine learning and knowledge discovery in databases, pp.* 273-289,*2020.*

[7] S. Minaee, M. Minaei and A.Abdolrashidi, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network," Sensors,Vol.21, No.9, 3046,2021.

[8] H.Y. Li, N.N. Wang, Y. Yu, X. Yang and X.B. Gao, "LBAN-IL: A novel method of high discriminative representation for facial expression recognition," Neurocomputing, vol. 432,pp.159–169 ,2021.

[9] A. Barman,P. Dutta, "Facial expression recognition using distance and shape signature features, pattern recognition letters," vol.145,pp. 254-261,2021.

[10] B.C.Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information, " Sensors,vol.18,No.2, 401, 2018.

[11] S.E. Kahou, X. Bouthillier, P. Lamblin,et al., "Emonets: Multimodal deep learning approaches for emotion recognition in video, " Journal on Multimodal User Interfaces, vol. 10,No.2,pp. 99–111,2016.

[12] F.J. Chang, A.T. Tran, T. Hassner, I. Masi, R. Nevatia and G. Medioni, "ExpNet: Landmark-Fee, Deep, 3D Facial Expressions,"*2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, IEEE,p.122-129,*2018*.

[13] R. Xu, J. Chen, J. Han, L. Tan and L. Xu, "Towards emotion-sensitive learning cognitive state analysis of big data in education: deep learning-based facial expression analysis using ordinal information," Computing,vol.102,pp.765–780,2020.

[14] H.J. Lee , D. Lee, "Study of Process-Focused Assessment Using an Algorithm for Facial Expression Recognition Based on a Deep Neural Network Model," electronics,vol.10,No.1, 54,2021.

[15] O. Arriaga, M. Valdenegro-Toro and P. Plöger, "Real-time convolutional neural networks for emotion and gender classification, " arXiv preprint arXiv:1710.07557, 2017.

[16] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features, " Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR, Kauai, HI, USA), pp. 1-9, 2001.