Research Article

Application of Convolutional Neural Network in Emotion Recognition of Ideological and Political Teachers in Colleges and Universities

Bo Gao [ORCID]

Xingjian College, Xijing University, Xi'an, Shaanxi 710123, China

Correspondence should be addressed to Bo Gao; 20120057@xijing.edu.cn

With the update of Internet technology and the development of we-media, ideological and political education in colleges and universities has been greatly impacted. Higher requirements are put forward for ideological and political teachers in colleges and universities, whose emotions seriously affect the quality and effect of teaching. Aiming at the problems of poor network generalization ability and large computation amount caused by many network parameters in the existing emotion recognition methods, a face emotion recognition method based on convolutional neural network is proposed. The network structure of nested Maxout multilayer perceptron layer is constructed by optimizing the convolutional neural model. Maxout can enhance the feature extraction capability of the convolutional layer of convolutional neural network. Meanwhile, Maxout performs linear combination of target features to select the most effective feature information. Then, the pretraining model is used for emotion recognition training. The strong perception ability of the model for facial features is retained by changing the important parameters. Simulation results demonstrate that this method has a higher recognition rate of face emotion and can effectively achieve accurate face emotion classification.

1. Introduction

China's education, especially higher education, has highlighted abundant problems with the rapid development in a new historical period [1]. In the new era, it is highly required to adhere to moral cultivation and put ideological and political work through the whole process of higher education teaching. At the highest level of the education system, university education is related to the future development of the country and the great rejuvenation of the Chinese nation. Therefore, more attention should be paid to the importance of ideological and political education [2]. Teachers, as engineers of human soul, shoulder the sacred mission of teaching and educating people. As a preacher, we should first understand the way and channel; that is, every word and action of teachers is the benchmark for students to learn [3]. Ideological and political teachers in colleges and universities are practitioners of ideological and political education, and

their emotions are the key factors affecting teaching quality and efficiency.

Emotion is an individual's subjective feeling and psychological behavior response to stimuli, and, also, has a strong dominant effect on individual psychological and physiological activities [4]. Both positive and negative emotional responses have a Sino-Arab influence on the normal functioning of the organism. With the further development of curriculum reform, the work intensity of teachers increases, and the teaching process becomes more complicated. Teachers are prone to emotional problems due to work pressure and complexity of irritable communication [5]. Improving teachers' emotional management ability is helpful to alleviate and solve teachers' emotional problems. Good emotional management ability can correctly guide positive emotions and maintain teachers' physical and mental health. At the same time, teachers' emotional reaction state has a direct impact on teaching quality and

efficiency. Teachers have excellent emotional management ability and can fully mobilize positive emotions to activate the classroom atmosphere, so as to ensure good results in education and teaching [6].

In the process of interpersonal communication, people usually judge the emotion of the other party according to the change of facial expression [7], so as to better communicate. Facial emotion is an extremely important way of language communication, and also an important mean of communication between people. Literature [8] divides emotions into six basic forms, including sadness, happiness, fear, disgust, surprise, and anger. Facial emotion recognition technology is a combination of physiology, psychology, image processing, machine vision, pattern recognition, and other research fields [9]. Moreover, it is a further development of face recognition, which mainly includes three steps of face image pretreatment, facial feature extraction and emotion recognition [10], as shown in Figure 1. This paper focuses on the direction of facial feature extraction for further research.

Convolution neural network (CNN) [11] is a branch of deep feedforward neural network, which has been widely applied in the field of image recognition. CNN is composed of one-dimensional, two-dimensional, and three-dimensional convolutional neural networks, which are applied to sequential data processing, image text recognition, and medical image and video data recognition, respectively. Literature [12] constructed a new 3D CNN convolution motion recognition method to obtain feature information from spatial and temporal dimensions. Literature [13] proposed a new postural convolutional neural network descriptor (P-CNN) for emotion recognition. Literature [14] proposed a method for driver emotion recognition based on convolutional neural network.

All the above studies on human emotion recognition are based on the extended convolutional neural network model require manual feature labelling. Its network generalization ability is poor, while the calculation amount is increased due to the network parameters. Besides, the ability of feature acquisition needs to be further improved. To solve these problems, this paper proposes a convolutional neural network based facial emotion recognition method. The innovations and contributions of this paper are listed below.

(1) Firstly, the convolutional neural network model is optimized, and then the nested Maxout multilayer perceptron network is constructed for improving the fitting ability of the algorithm and improving the recognition accuracy of the model.

(2) The target features are linearly combined by nested Maxout to select the most effective feature information. The experimental results show that the proposed method achieves good effect on the accuracy of teachers' emotion recognition, and can accurately classify faces' emotions.

The chapter structure of this paper is as follows. The recommendation algorithm proposed in this paper is
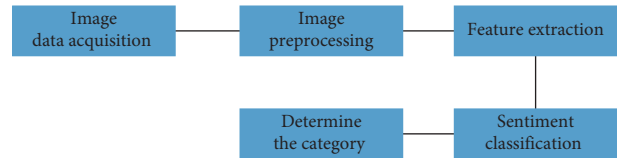

FIGURE 1: Main steps of facial expression recognition.

described in the next s. The convolutional neural network structure is constructed in section three. Section four focuses on the neural network training. Section five is experiment and analysis. Section six is the conclusion.

2. The Recommendation Algorithm Proposed in This Paper

2.1. Structural Characteristics of Convolutional Neural Network. Network structures such as multilayer perceptron, convolution kernel, pooling layer, local connection, and full-time sharing are widely used in convolutional neural network structures. These applications greatly reduce the time and space complexity of neural network. In addition, it greatly reduces the weight parameters of network structure, which is also beneficial to the training of neural network.

2.1.1. Local Connection. Local connections are also called sparse connections. Inspired by the visual neural structure in biology, neurons in the visual cortex receive local information (that is, these neurons only respond to stimuli in certain areas). The spatial relation of image pixels is strongly correlated with the close pixels; otherwise, the correlation is weak. As a result, the neuron receives only the local receptive field it is responsible for, and does not need to perceive all pixels. The local information of perception is then integrated into global perception by local information fusion of the next layer. Local connection can greatly reduce the number of weights between layers of convolutional neural network and carry out feature dimension reduction. Then, effective features are screened for neural network learning and training to improve the learning efficiency of the model.

2.1.2. Weight Sharing. Weight sharing means that the same convolution kernel is used to process the whole input image. Features extracted locally are the same as those extracted in other parts, and the same learning features can be used in other locations. The weight sharing of convolutional neural network reduces feature dimension and parameter number. Also, the time and space complexity of neural network are reduced.

2.1.3. Multilayer Convolution Kernel. After the first convolutional layer of the convolutional neural network carries out the convolution operation, the feature graph obtained by the convolutional layer is some shallow features of the image, such as edge information and line outline, in addition to other information. For image recognition, deep features are needed, and shallow features cannot fully express the

semantic information of the image. One convolution kernel can only obtain the same feature graph. To obtain deeper features, multilayer convolution kernel is required to extract feature information and form feature maps of various information.

In the field of image recognition, the feature hierarchy of input image is born. As shown in Figure 2, you start with the original input pixels and go up to simple lines and textures made up of pixels. Then, the lines and textures form patterns, and finally the individual patterns form objects in the image. In the whole process, the shallow features are found through the original input, and then the shallow features are further mined to find the middle features, and the last step is to obtain the deep features. It is impossible to find deep features directly from raw inputs. In short, single-layer convolution usually acquires shallow features, and it is possible to acquire deeper features by increasing the number of convolution layers.

2.1.4. Principle of Convolution Process. The convolutional layer carries out the convolution operation on the image, and the obtained feature map contains the structural features of the original image, and the deep-seated features can better express the essential meaning information of the image. The convolution of the function is defined as follows. For two continuous integrable functions $h(i)$ and $a(i)$ on R, their convolution $b(i)$ is as follows:.

$$b(i) = \int_{-\infty}^{\infty} h(\tau)a(i - \tau)\mathrm{d}\tau, \tag{1}$$

where the convolution of $h(i)$ on $a(i)$ is denoted as $h(i) * a(i)$, representing the integral of the product of and $a(\alpha - i)$ in the domain of definition. α is the independent variable of the convolution function $b(i)$, which is the position of the convolution.

The convolution calculation process is to transform the picture into a data matrix, and the wandering window is a convolution kernel matrix. After the convolution processing of $M * M$ convolution kernel for an $N * N$ image, the feature graph of $(N - M + 1) * (N - M + 1)$ will be obtained.

2.2. Softmax Classifier. The promotion and application of Logistic regression model [15] formed Softmax classifier to solve the problem of multiple classification. The optimized convolutional neural network model in this paper uses Softmax to classify behaviors. Assume that abnormal behaviors are divided into z and classified. There are w video sequences of sample data. Suppose the convolutional neural network training dataset is N.

$$N = \left\{ \left(i^{(1)}, j^{(1)} \right), \left(i^{(2)}, j^{(2)} \right), \dots, \left(i^{(w)}, j^{(w)} \right) \right\}, \tag{2}$$

where $i^{(x)}$ is the x input sample. $j^{(x)}$ is the behavior label of sample x, $j^{(x)} \in \{1, 2, \dots, z\}$.

For each input $i^{(x)}$, Softmax classifier calculates the probability for each class. The calculation equation is as follows:

$$U(j = yi)j = 1, 2, \dots, z. \tag{3}$$

From the vector point of view, the equation for calculating the function is as follows:

$$h\left(i^{(x)}|\theta\right) = \begin{bmatrix} u\left(j^{(x)} = 1|i^{(x)}, \theta\right) \\ u\left(j^{(x)} = 2|i^{(x)}, \theta\right) \\ \vdots \\ u\left(j^{(x)} = z|i^{(x)}, \theta\right) \end{bmatrix}$$

$$= \frac{1}{\sum_{y=1}^{z} g_y^{\theta_y^T i^{(x)}}} \begin{bmatrix} e^{\theta_1^T i^{(x)}} \\ e^{\theta_2^T i^{(x)}} \\ \vdots \\ e^{\theta_z^T i^{(x)}} \end{bmatrix}. \tag{4}$$

In the equation, θ represents neural network parameters. So, there are z behaviors, and each behavior has a probability value. The value of probability ranges as follows [0,1], and the probability sum of z abnormal behaviors is 1. The output of the neural network corresponds to the probability of the behavior and that probability corresponds to the label of the behavior.

During neural network training, Softmax is used for behavior classification, and the loss function is calculated as follows:

$$L(\theta) = -\frac{1}{w} \sum_{x=1}^{w} \sum_{y=1}^{z} 1\left\{y^{(x)} = j\right\} \log \frac{e_j^{\theta} x^{(i)}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}, \tag{5}$$

where $1\{y^{(x)} = j\}$ represents an exponential function. When $y^{(x)}$ is equal to j, the output is 1. Otherwise, the output is 0, and its output is the label matrix of abnormal behavior.

In general, gradient descent algorithm is used to calculate the loss function in the process of backpropagation, and the calculation equation is as follows:

$$\frac{\partial L(\theta)}{\partial \theta} = -\frac{1}{w} i^{(x)} \left[1\left\{j^{(x)} = y\right\} - u\left(j^{(x)} = y \mid i^{(x)}; \theta\right) \right]. \tag{6}$$

Equation (6) is used to obtain the gradient of the loss function to the weight function, and the gradient is used to guide the adjustment of neural network model parameters until the end of neural network training and the optimal weight parameters are obtained.

3. Construct Convolutional Neural Network Structure

Traditional CNN uses single-layer linear convolution in the convolution layer, which does not perform well in the extraction of nonlinear features and abstract features hidden in complex images. The activation function has strong fitting ability and can fit all characteristic patterns when the number of neurons is sufficient. Therefore, the nested
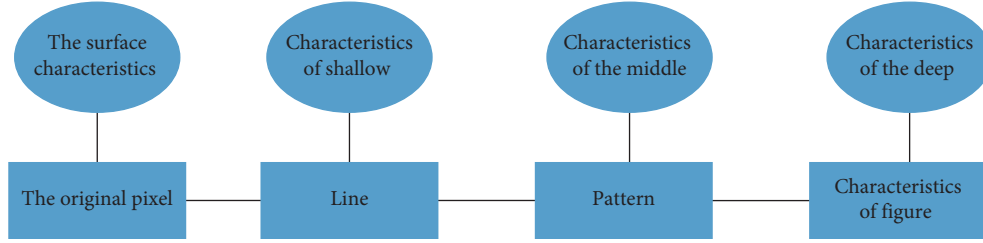
FIGURE 2: Schematic diagram of feature extraction process.

Maxout MLP (Multilayer Perception) layer [16] is combined with the activation function to improve the fitting ability of the algorithm and the recognition accuracy of the model.

3.1. Determination of Nesting Layers.

The number of linear regions in neural networks with nested Maxout layers increases as the number of Maxout layers increases. In addition, the number of linear regions in ReLU and Maxout networks increases exponentially with the number of layers. Maxout networks tend to overfit training datasets in the absence of model regularization. It is attributed to the fact that Maxout network can recognize the most valuable input information in the training process and is easy to carry out feature coadaptation.

The method in this paper was tested on a dataset using a different number of Maxout layer fragments, as shown in Figure 3. The test results of combining Maxout fragments with Maxout layer and Batch Normalization (BN) [17] layer fragments show that the nested model has reached saturation state when Maxout fragment is 5. As shown in Figure 3, the number of layers 5 is the best choice.

3.2. Selection of Pooling Layer.

In general, researchers will select the largest pooling layer for sampling, which is more representative in extracting features by using average pooling pools effective features across all pooling layers. The irrelevant feature information in the input image can be suppressed by average pooling and discarded by maximum merging. The average pool is an extension of the global average pool, where the model tries to extract information from each local patch to facilitate abstraction into feature maps. The nested structure can extract abstract representative information from each part, making more distinguishable information embedded in the feature map. Spatial average pooling is used in each pooling layer to aggregate local spatial information. In the CIFAR-10 dataset without data expansion, the comparison results of test error rates of the maximum and average pooling layer are shown in Table 1.

3.3. Building a Nesting Layer.

Convolution layer of nested multilayer Maxout network is constructed. That is, Maxout MLP is used to extract features based on nested network structure, and the constructed convolutional neural network
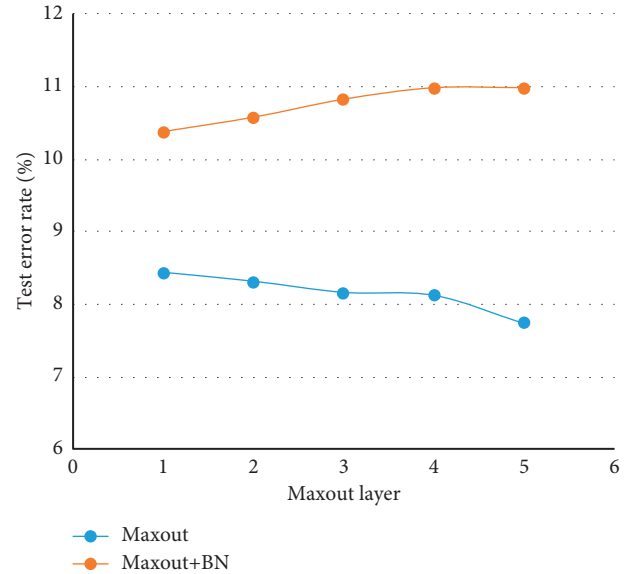

FIGURE 3: Different number of Maxout layer test results.

TABLE 1: Comparison of maximum and average pooling layer test error rates.

The adopted pooling layer	Test error rate/%
Max pooling	8.75
Avg pooling	7.81

model uses batch standardization to reduce saturation and pressure difference to prevent overfitting. In addition, the basic features obtained by mean pool aggregation Maxout MLP are applied across all pool layer to increase the robustness of object space transformation as follows:

$$h_{x,y,z} = \max_{w \in [1,t]} \left(m_{z_w}^T i_{x,y} + d_{z_w} \right), \qquad (7)$$

where (x, y) is the position of pixels in the feature graph. $i_{x,y}$ is an input block centered on pixel (x, y). z_w is the channel $h_{x,y,z}$ for indexed feature mapping. T is the number of MLP layers. From another perspective, the Maxout unit is equivalent to the cross-channel maximum pooling layer on the convolution layer. Cross-channel maximum pooling layer selects the maximum output to be entered for the next layer. Maxout cells help solve the problem of fading gradients because gradients flow through each maximum cell.

The feature mapping in the nested Maxout MLP layer module is calculated as follows:

$$h^1_{x,y,t_1} = BN\left(\left(m^1_{t_1}\right)^T i_{x,y} + d^1_{t_y}\right),$$

$$h^2_{x,y,t_2} = \max_{w \in [1,z]}\left(BN\left(\left(m^2_{t_w}\right)^T h^1_{x,y} + d^2_{t_w}\right)\right),$$ (8)

$$h^3_{x,y,t_3} = \max_{w \in [1,z]}\left(BN\left(\left(m^3_{t_w}\right)^T h^2_{x,y} + d^3_{t_w}\right)\right),$$

where $BN(\cdot)$ represents the batch normalization layer. (x, y) is the position of pixels in the feature graph. $i_{x,y}$ is an input block centered on pixel (x, y). z_t is the serial number of each channel in the feature graph. t is the number of layers of nested Maxout MLP. The batch standardization layer can be applied before activating the function. In this case, nonlinear elements tend to produce activation with a stable distribution, reducing saturation. As shown in Figure 4, the convolutional layer structure diagram of nested Maxout layer is constructed.

3.4. Building a Nested Maxout Layer Convolutional Neural Network Model. By superposing the convolutional layer model of four nested Maxout layers, the whole structure of convolutional neural network with nested Maxout MLP layers is formed.

The network structure of the nested Maxout MLP layer is equivalent to a cascading cross-channel parameter pool and a cross-channel maximum pool on the convolutional layer. Nested structures can combine feature maps linearly and select the combination of the most effective information to output to the next layer. The nested structure reduces saturation by applying batch normalization and can encode the information in the activation patterns of paths or Maxout fragments, thus enhancing the discrimination ability of the deep architecture of the convolutional neural network.

4. Train the Neural Network

The training process of neural network model adopts error backpropagation algorithm and is divided into forward propagation stage and backpropagation stage. In the forward propagation stage, each hidden layer of the neural network receives the output of the previous layer, and the output of this layer is calculated by activating the activation function. In the backpropagation stage, the loss function is used to calculate the output error of the neural network, meanwhile, the error of each hidden layer of the neural network is calculated layer by layer. The error of each hidden layer is used as the updating basis of the weight parameters of the previous hidden layer. The training steps of neural network algorithm are shown in Figure 5.

Step 1. Randomly initialize the ownership value and threshold of the network. The value range is $(-1, 1)$.

Step 2. For the training samples $(i_x, j_x,)$, the actual network output is calculated as follows:

$$\widehat{j}^x_y = h\left(\beta_y - \theta_y\right).$$ (9)

In the equation, $f(\cdot)$ represents the activation function Sigmoid function. θ_y represents the threshold of the y neuron in the output layer of the neural network. β_y represents the input of the y neuron in the output layer of the neural network:

$$\beta_j = \sum_{x=1}^{t} m_{x,y} i_x,$$ (10)

where $m_{x,y}$ represents the weight between the x neuron in the hidden layer of the neural network, and the y neuron in the output layer of the neural network.

Step 3. Calculate the mean square error of convolutional neural network on (i_x, j_x). The calculation equation is as follows:

$$G = \frac{1}{2} \sum_{x=1}^{w} \left(\widehat{j}^x_y - j^x_y\right)^2,$$ (11)

where $\widehat{j}^x_y$ represents the actual output of the convolutional neural network. j^x_y represents the expected output of the convolutional neural network.

Step 4. Check whether the conditions are met, that is, whether the error is less than the minimum value allowed by the learning error or the learning time reaches the set minimum number. If the conditions are not met, the weights of the convolutional neural network are updated, and the weights and thresholds of the neural network are adjusted according to the gradient direction of the target. Assume that the learning rate of neural network training process is η, and the weight update calculation equation of neural network is as follows:

$$\Delta m_{xy} = -\eta \frac{\partial G}{\partial m_{xy}}$$

$$= \eta \widehat{j}^x_y\left(1 - \widehat{j}^x_j\right)\left(j^x_y - \widehat{j}^x_j\right) i_x.$$ (12)

Step 5. Repeat Step 2 to Step 4 until the end condition is met. That is, the neural network training is complete, and the weights and thresholds of the neural network are fixed.

5. Experimental Results and Data Analysis

In this paper, the GPU version of PyTorch framework is adopted, and the hardware platform is Ubuntu 20.04.2 with dual-core Intel 4.4ghz CPU, Tesla K80 GPU, 2 TB hard disk memory, and 12 GB running memory. Because there is no public dataset related to ideological and political teachers' emotions, the two widely used datasets are used to verify the validity of the algorithm, such as CK + dataset and Oulu-CASIA dataset. The input images of the two tasks were normalized to $112 * 112$ pixels after face detection and matching using multi-task convolutional neural network. The stochastic gradient descent of the driving quantity was used as the
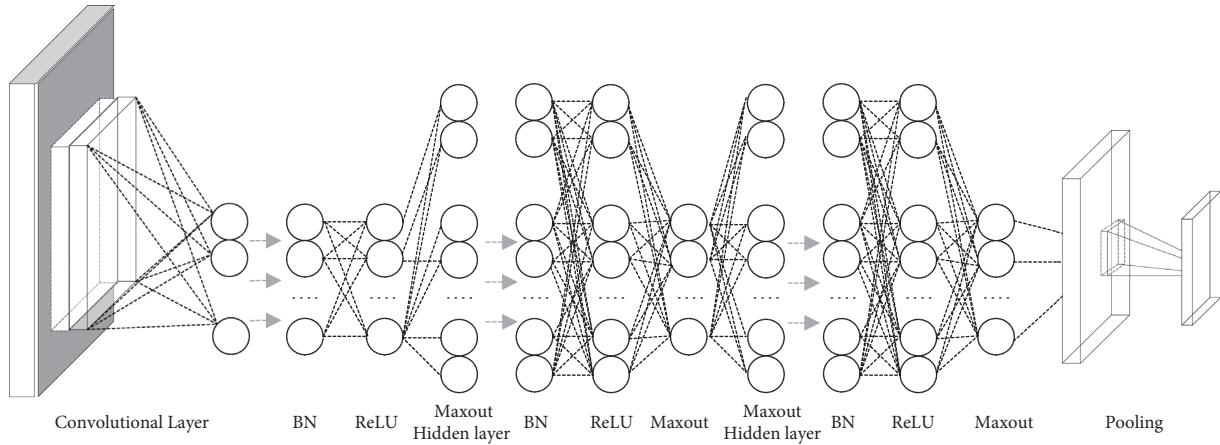
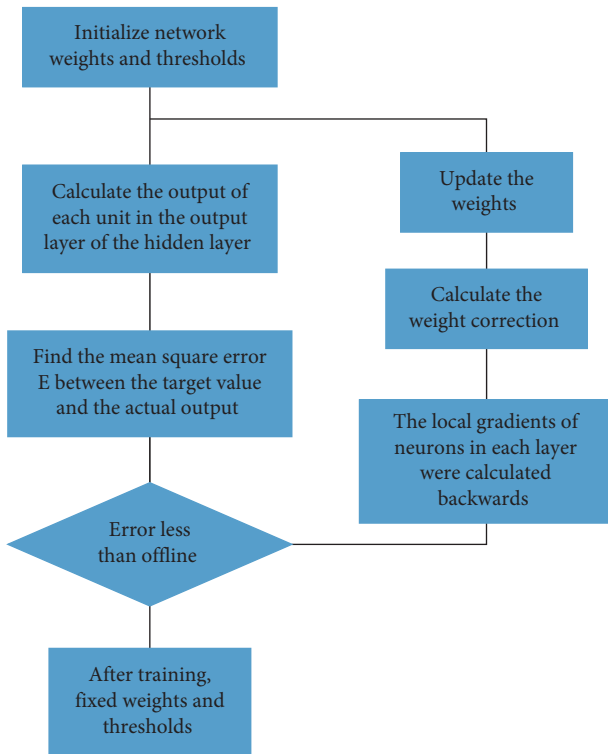FIGURE 4: The CNN layer structure diagram of nested Maxout layer.

FIGURE 5: The training steps of proposed algorithm.

optimizer in both training stages, and the momentum was set as 0.9, and the weight attenuation term was set as 0.0005. Random horizontal flip was used for data enhancement. In the first stage, CASIA-WebFace is used to train the face recognition model. ArcFace was selected as the loss function, and the batch size was set to 256. A total of 70 epochs were trained in the network, and the initial learning rate was 0.1. Since the 40th epoch, the learning rate of every 10 epochs has decreased to 1/10 of that of the previous epoch. After the face recognition model training, the part before the full connection layer was used for the second stage emotion recognition training. At this time, the batch size was set as 32, the initial learning rate was set as 0.01, and the learning rate was reduced by 1/10 of the current value for every 5 epochs, and a total of 25 epochs were trained.

5.1. CK+ Dataset. CK+ is a dataset collected in a laboratory setting, containing 593 video sequences from 123 subjects. Six basic emotions were classified in the experiment, and only 309 sequences were selected from 106 subjects. Then, 927 images were extracted from the last 3 frames of each sequence where the emotional intensity peaked. Finally, the selected images are divided into 10 subsets according to the ascending order of person ID, and the identity-independent tenfold cross verification is carried out. The average accuracy of cross-validation is shown in Table 2. As can be indicated from Table 1, the proposed method achieves 98.75% recognition rate. Figure 6 shows the confusion matrix on CK+. It can be seen that the algorithm has the best recognition effect on disgust, happiness, and sadness, while anger and fear have fewer samples and are relatively difficult to identify.

5.2. Oulu-CASIA Dataset. Oulu-CASIA is a dataset collected in a laboratory setting. It includes sequences of 2880 images from 80 subjects and is labeled with six categories of basic emotion labels. 480 image sequences taken under normal lighting conditions were selected. Similar to CK+ database, the last 3 peak frames in each sequence are selected to form a total of 1440 images, and each emotion has the same number of images. Then, identity-independent is done tenfold cross authentication. Table 3 shows the recognition rates of the six basic emotions on Oulu-CASIA.

It can be seen that the proposed method achieves comparable results with advanced algorithms. The accuracy of the algorithm is slightly lower than fine tuning because the pretrained face recognition network is trained by using Internet images. Oulu-CASIA's images include a variety of light conditions in a laboratory setting. This difference in image distribution makes the pretrained face model have weak perception ability for such images, thus resulting in poor performance. Figure 7 is the confusion matrix on Oulu-CASIA. It is indicated that the algorithm has the best performance in recognizing happiness and surprise, while its performance in recognizing fear and disgust is relatively weak.

Table 2: The overall accuracy on CK + database.

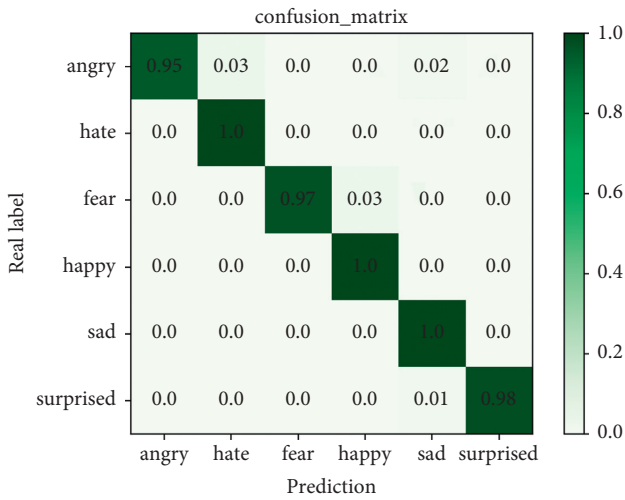| Methods	Precision (%)
Literature [18]	97.14
Literature [19]	97.29
Literature [20]	98.41
Literature [21]	98.32
Literature [22]	98.62
The proposed method	**98.75**

Figure 6: The confusion matrix on the CK+ dataset.

Table 3: The overall accuracy on Oulu-CASIA database.

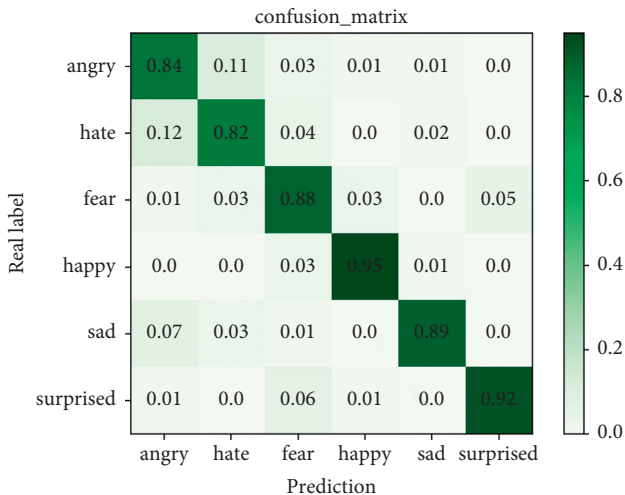| Methods	Precision/%
Literature [18]	81.35
Literature [19]	84.48
Literature [20]	87.64
Literature [23]	88.12
Literature [24]	88.24
Literature [25]	86.47
The proposed method	**88.31**

Figure 7: The confusion matrix on the Oulu-CASIA dataset.

6. Conclusion

In the new era, strengthening ideological and political education in colleges and universities has far-reaching practical significance. Ideological and political education not only can promote the comprehensive and harmonious development of college students, but also can be the inevitable requirement of building a harmonious society. With the development of curriculum reform, colleges and universities put forward higher requirements for ideological and political teachers. Since ideological and political teachers serve as the executor of ideological and political education, personal emotions of those people directly affect the effect and quality of teaching. Therefore, ideological and political teachers should pay attention to the change of their own emotions, trying to improve their own emotional management ability. This paper proposes the application of convolutional neural network in emotion recognition of ideological and political teachers in universities. Through the network structure of nested Maxout MLP layer, the ability of neural network to extract nonlinear features and abstract features hidden in complex images is improved. Using the activation function, ReLU in the nested layer can improve the performance of neural networks and feature patterns. Nested structures use batch normalization to desaturate. Simultaneously, the information in activation mode of path or Maxout fragment is encoded to enhance the discrimination ability of deep architecture of convolutional neural network. As the ability of extracting facial features is basically retained, the model's processing ability for the diversity of facial expression images in real world environment is enhanced, the performance is improved more obviously, and the recognition rate is higher. In the future, the author will analyze the performance of the algorithm in this paper and establish a public facial expression recognition database of ideological and political teachers in universities.

Abbreviations

CNN: Convolution neural network
P-CNN: Postural convolutional neural network
MLP: Multilayer Perception.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no competing interests.

References

[1] J. Huang, "Successes and challenges: online teaching and learning of chemistry in higher education in China in the time of COVID-19," *Journal of Chemical Education*, vol. 97, no. 9, pp. 2810–2814, 2020.

[2] J. Zhang, X. Hu, Z. Ning et al., "Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks," *IEEE Internet of Things Journal*, Date of Publication, vol. 5, no. 4, pp. 2633–2645, 2018.

[3] L. A. Shepard, K. L. Davidson, and R. Bowman, "How Middle-School Mathematics Teachers Use Interim and Benchmark Assessment Data. CRESST Report 807," *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*, Oxford University, Oxford, UK, 2011.

[4] H. D. Critchley and S. N. Garfinkel, "Interoception and emotion," *Current opinion in psychology*, vol. 17, pp. 7–14, 2017.

[5] K. Y. Wong, T. Sulaiman, A. Ibrahim, A. G. K Mohd, O Hassan, and W. M. W Jaafar, "Secondary School Teachers Psychological Status and Competencies in E-Teaching during Covid-19," *Heliyon*, vol. 7, Article ID e08238, 2021.

[6] P. A. Jennings, *Mindfulness for Teachers: Simple Skills for Peace and Productivity in the Classroom (The Norton Series on the Social Neuroscience of education)*, WW Norton & Company, NY, USA, 2015.

[7] M. G. Calvo and L. Nummenmaa, "Perceptual and affective mechanisms in facial expression recognition: an integrative review," *Cognition & Emotion*, vol. 30, no. 6, pp. 1081–1106, 2016.

[8] Y. Susanto, A. G. Livingstone, B. C. Ng, E. Cambria, and E. Cambria, "The hourglass model revisited," *IEEE Intelligent Systems*, vol. 35, no. 5, pp. 96–102, 2020.

[9] Z. Ning, X. Hu, Z. Chen et al., "A cooperative quality-aware service access system for social Internet of vehicles," *IEEE Internet of Things Journal*, Date of Publication, vol. 5, no. 4, pp. 2506–2517, 2018.

[10] K. Dharavath, F. A. Talukdar, and R. H. Laskar, "Improving face recognition rate with image preprocessing," *Indian Journal of Science and Technology*, vol. 7, no. 8, pp. 1170–1175, 2014.

[11] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network[C]//2017 international conference on engineering and technology (ICET)," *IEEE*, pp. 1–6, 2017.

[12] H. Peng, B. Hu, Q. Shi et al., "Removal of ocular artifacts in EEG-an improved approach combining DWT and ANC for portable Applications," *IEEE journal of biomedical and health informatics*, Date of Publication, vol. 17, no. 3, pp. 600–607, 2013.

[13] M. Khokhlova, C. Migniot, and A. Dipanda, "3D Point Cloud Descriptor for Posture Recognition," in *Proceedings of the International Conference on Computer Vision Theory and Applications*, pp. 161–168, Dalian, China, January 2018.

[14] S. Yan, Y. Teng, J. S. Smith, and B Zhang, "Driver Behavior Recognition Based on Deep Convolutional Neural networks," in *Proceedings of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 636–641, IEEE, Changsha, China, August 2016.

[15] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic regression*, John Wiley & Sons, NJ, USA, 2013.

[16] J. M. Pérez-Rúa, V. Vielzeuf, S. Pateux, M Baccouche, and F Jurie, "Mfas: Multimodal Fusion Architecture search," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6966–6975, Long Beach, CA, USA, June 2019.

[17] X. Hu, J. Cheng, M. Zhou et al., "Emotion-Aware cognitive system in multi-channel cognitive radio ad hoc networks," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 180–187, 2018.

[18] H. Jung, S. Lee, S. Park, I Lee, C Ahn, and J Kim, "Deep temporal appearance-geometry network for facial expression recognition," 2015, https://arxiv.org/abs/1503.01532.

[19] Z. Yu, Q. Liu, and G. Liu, "Deeper cascaded peak-piloted network for weak expression recognition," *The Visual Computer*, vol. 34, no. 12, pp. 1691–1699, 2018.

[20] W. Wei and K. Qiao, "Jakub nowak, marcinkorytkowski rafa? Scherer, marcin Wo?niak, accurate and fast url phishing detector: a convolutional neural network approach," *Computer Networks*, vol. 178, Article ID 107275, 2020.

[21] Y. Li, J. Liu, Z. Tang, and B. Lei, "Deep spatial-temporal feature fusion from adaptive dynamic functional connectivity for MCI identification," *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2818–2830, 2020.

[22] W. Wei, H. Song, W. Li, P. Shen, and A. Vasilakos, "Gradient-driven parking navigation using a continuous information potential field based on wireless sensor network," *Information Sciences*, vol. 408, no. 2, pp. 100–114, 2017.

[23] H. Yang, U. Ciftci, and L. Yin, "Facial Expression Recognition by De-expression Residue learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2168–2177, Salt Lake City, UT, USA, June 2018.

[24] X. Liu, B. V. K. Vijaya Kumar, P. Jia, and J. You, "Hard negative generation for identity-disentangled facial expression recognition," *Pattern Recognition*, vol. 88, pp. 1–12, 2019.

[25] D. Gera and S. Balasubramanian, "Landmark guidance independent spatio-channel attention and complementary context information based facial expression recognition," *Pattern Recognition Letters*, vol. 145, pp. 58–66, 2021.