Feature Extraction Method of Students' Ideological and Political Learning Behavior Based on Convolutional Neural Network

Hong-bing Jiang(✉)

School of Road Bridge and Architecture, Chongqing Vocational College of Transportation, Chongqing 402247, China

Abstract. In order to improve the accuracy of feature extraction of students' Ideological and political learning behavior, a method of feature extraction of students' Ideological and political learning behavior based on convolutional neural network is proposed. The image of students' Ideological and political learning behavior is obtained and stored, and the stored image is corrected. On the basis of image correction, the similarity measurement of students' image spatial structure information details and the representation of image spatial structure information details are used to extract the characteristics of students' Ideological and political learning behavior based on convolutional neural network. The experimental results show that the method based on convolution neural network not only improves the accuracy of feature extraction, but also reduces the time of feature extraction.

Keywords: Convolution neural network · Learning behavior · Feature extraction

1 Introduction

As the main body, learners play an important role in teaching activities. As a reflection of students' physical and psychological conditions, learning state has become a hot research content in the field of education. As one aspect of learning state, fatigue state can better reflect the performance of students in the classroom. The research on it is helpful to understand students and teaching evaluation. Therefore, it is necessary to study the feature extraction of students' learning behavior.

Facial features are an effective representation of students' learning characteristics, and studies have shown that eyes are most closely related to fatigue. The detection steps in current research generally include face detection, eye location and detection, feature extraction, eyes and fatigue status judgment. Among them, face and eye detection and state judgment are the core issues. However, the main problem existing in the existing research is that the positioning of the eyes is easily affected by the external environment, and the judgment of the state requires artificial definition and extraction of eye features. Only by solving these key problems, can the judgment of students' ideological and political learning behavior be effectively realized. For this reason, aiming at the problems in traditional methods, a feature extraction method of students' ideological and political

W. Fu et al. (Eds.): eLEOT 2021, LNICST 390, pp. 514–526, 2021.
https://doi.org/10.1007/978-3-030-84386-1_42

learning behavior based on convolutional neural network is proposed. Convolutional neural network is a construction that imitates the biological visual perception mechanism, which has advantages when processing big data such as images. It is different from traditional neural networks in that it uses convolution operations instead of multiplication operations. The connection between the convolutional layers in the convolutional neural network is called sparse connection, that is, compared with the full connection in the feedforward neural network, a certain neuron in the convolutional layer is only connected to a part of its adjacent layer, rather than all neurons. In addition, the convolutional layer and pooling layer in the convolutional neural network can respond to the translation invariance of input features, that is, it can identify similar features located at different positions in space. Therefore, this paper uses the convolutional neural network to study the students' thinking behavior characteristic extraction method. First, store the acquired student's thinking to study behavior, correction processing, according to processing results, information details in the image Similarity metrics, characterize the details of image space structure information, based on this, extract students' thinking and regulatory learning behavior, and finally verify the effectiveness of this method through simulation experiments.

2 Image Acquisition and Storage of Students' Ideological and Political Learning Behavior

Training, detection and testing process are inseparable from experimental data, image acquisition and storage is to collect experimental data. With the help of camera, students' activity videos in class are captured dynamically, and the video stream is converted into images by opencv according to the actual situation, and stored in the specified folder, so as to be used as the input data of later face detection experiment.

In order to understand human behavior, we need to detect human objects through recognition system. Firstly, the human and video frames in the video sequence are decomposed. The time frame method can be used to obtain the moving region of the image by comparing the pixel difference corresponding to two or three images in the adjacent time of the video sequence in the recognition system, and the calculation formula is as follows:

$$E(m,n) = \sum_{i=1}^{o} \kappa d_i B \tag{1}$$

Assuming that the pixel value of the o-th frame image in the video at the point (m, n) is represented as d_i, and the difference between the two frames of image is represented by B, and no directional analysis is performed in this calculation.

According to the above process, the images of students' ideological and political learning behavior are acquired and stored, which provides a basic basis for the extraction of characteristics of students' ideological and political learning behavior.

3 Image Correction Processing

The image data set of students' ideological and political learning behavior can be obtained by converting the video, but in order to make the characteristics more obvious,

the image is generally preprocessed before the experiment. As the image acquisition process is affected by many factors, the accuracy of the recognition result is low. To this end, the image of students' ideological and political learning behavior is corrected, and the correction objective function is set [1]. Use a suitable color template to adjust the processed color to get a good correction effect. In terms of color research, the HSL color model is more suitable for human expression; the HSL color model obtains different colors by changing the hue, saturation and lightness [2]. The model can basically contain all the colors that human vision can perceive, and the color description can correspond to H, S, and l. The method of converting RGB to HSL is as follows: R, G, B represent the red, green, and blue of a specific color in turn, and the value varies between 0 and 1. The detailed calculation formula is as follows:

$$H=\begin{cases} 0, if\, \lambda_{\max}=\lambda_{\min} \\ 60^\circ \frac{G-B}{\lambda_{\max}-\lambda_{\min}}+0^\circ, if\, \lambda_{\max}=R, G>B \\ 60^\circ \frac{G-B}{\lambda_{\max}-\lambda_{\min}}+360^\circ, if\, \lambda_{\max}=R, G<B \\ 60^\circ \frac{B-R}{\lambda_{\max}-\lambda_{\min}}+120^\circ, if\, \lambda_{\max}=G \\ 60^\circ \frac{R-G}{\lambda_{\max}-\lambda_{\min}}+240^\circ, if\, \lambda_{\max}=B \end{cases} \tag{2}$$

In formula (2), $\lambda_{\max}$ represents the largest value in the range of 0–1, and $\lambda_{\min}$ represents the smallest.

On the basis of the above analysis, the image is converted to HSL color space, and the color disharmony coefficient is calculated by the following formula:

$$z[P,(n,\varepsilon)]=\sum_{j=P}\left|H(q)-S_{M_{n(e)}}(q)\right|\eta_A(q) \tag{3}$$

In formula (3), ε represents the rotation angle description parameter, P represents all the color points describing the students in the image, $H(q)$ represents the color point describing the color value of q, $S_{M_{n(e)}}(q)$ represents the boundary color value description parameter, and $\eta_A(q)$ represents the area ratio of the q color point.

On this basis, the harmonious degree of students' images is fully analyzed, and the influencing factors can be obtained from the following formula:

$$W=\sum_{m-1}^{i=1}\frac{D_i^{\mathrm{MC}}}{D_{i+1}^{\mathrm{MC}}} \tag{4}$$

In formula (4), D_i^{MC} and D_{i+1}^{MC} respectively represent the degree of color effect of the image.

At the same time, individual selection and selection operation is a method of genetic algorithm to evaluate individual adaptability, and it is also the main way of genetic algorithm [3] to realize group gene transmission. Among them, the selection algorithm is obtained by roulette wheel selection, and the selection probability of each individual in the group is calculated by the following formula:

$$p_1(i) = \frac{F_w}{\sum_{j=1}^{N} F_w} \tag{5}$$

In genetic algorithm, fork is a key search operator [4]. It simulates the process of gene recombination in nature, passing good genes to the next generation, and generating better genetic structure. Under the condition of local convergence, mutation can expand the new search space and ensure the diversity of the population. The cross probability function and variation probability function can be obtained by the following formula:

$$p_c = \begin{cases} \frac{k_1(f_{\max} - f')}{(f_{\max} - \overline{f})}, f^{\circ} \geq \overline{f} \\ k_2, f < \overline{f} \end{cases} \tag{6}$$

$$p_{\mathrm{m}} = \begin{cases} \frac{k_3(f_{mux} - f)}{(f_{\max} - \overline{f})}, f \geq \overline{f} \\ k_4, f < \overline{f} \end{cases} \tag{7}$$

In formula (6) and formula (7), $f_{\max}$ represents the maximum fitness value of the description group, $\overline{f}$ represents the average fitness value used to describe the population of different generations, and f represents the individual with the larger fitness value among the two individuals to be crossed.

Continue to iterate the above calculation process until the fitness calculation result is obtained and the establishment of the correction target is completed.

4 Similarity Measurement of Spatial Structure Information of Students' Images

On the basis of the above-mentioned image correction of students' Ideological and political learning behavior, the similarity of information details in the image is measured. The specific steps are as follows:

Step1: The image set to be characterized is outputted with the transformation parameter θ through the porous convolutional neural network [5] structure;
Step2: According to the above parameter θ, the inverse coordinate mapping is realized after affine transformation, and the sampling network T before the input and output images is obtained;
Step3: The output result of sampling network T is processed by bilinear difference technology to obtain transformed image b.

Assuming that (x_t, y_t) corresponds to the pixel coordinates in input M, and (x_b, y_b) corresponds to the pixel coordinates in output N [6], the transformation parameter θ in step 2 is reversed coordinate mapping, and the similarity measurement process is:

$$(x_t, y_t) = TG * N(x_b, y_b) \tag{8}$$

In formula (8), TG represents affine transformation, that is, affine transformation is performed on the grid, and the transformed network is filled with the pixel values of corresponding coordinate points in the original image to obtain the real pixel value. The image expression after similarity measurement is as follows:

$$V_z^q = \sum_d a \sum_e y(x_i - m) \tag{9}$$

In formula (9), V_z^q represents the pixel value of the original image coordinate point, $\sum_d a$ represents the real pixel value of the image, $\sum_e y$ represents the feature parameters of the current layer and the previous layer in the convolutional layer, and x_i represents the convolution kernel of the feature map [7], m is the image parameter after transformation.

Through the above derivation, the similarity measurement of the spatial structure information of the image to be represented is realized, and the probability of the salient region with higher activation value is increased, providing a basis for the detailed characterization of the spatial structure information of the image spatial structure information of the students' ideological and political learning behavior.

According to the similarity measurement results of the image spatial structure information details of the students' ideological and political learning behavior, the image spatial structure information details are encoded. For shape feature extraction, the image to be represented is regarded as a whole, the image features in the area pixels are counted, and the area shape features are described. The extraction formula is as follows:

$$Z_{nn} = \frac{n+1}{\varsigma} \int_x^z a \tag{10}$$

In formula (10), Z_{nn} represents the eccentricity of the shape, $\frac{n+1}{\varsigma}$ represents the shape feature of the image, $n + 1$ represents the image feature extraction parameter, and $\int_x^z a$ represents the image region pixel.

5 Image Spatial Structure Information Detail Representation

On the basis of the similarity measurement and coding of the image spatial structure information details of the students' Ideological and political learning behavior, the image spatial structure information details are represented. Because the similarity measure of image spatial structure information details and the coding process of image spatial structure information details contain a lot of redundant information [8], the representation of image spatial structure information details is affected. Therefore, the purpose of removing the redundant information of students' Ideological and political learning behavior image is to restore the original image information from the image corrupted by noise. The calculation process is as follows:

$$g(x, y) = d(i, j) + b(i, j) \tag{11}$$

In formula (11), $g(x, y)$ represents the actual image, $d(i, j)$ represents the noise-free image [9], and $b(i, j)$ represents the added noise information.

After the actual image is processed by noise, it degenerates into a noisy image. Therefore, the porous convolutional neural network is used to deeply consider the relationship between the noisy image and the denoised image. The structure of the porous convolutional neural network is shown in the following figure (Fig. 1):

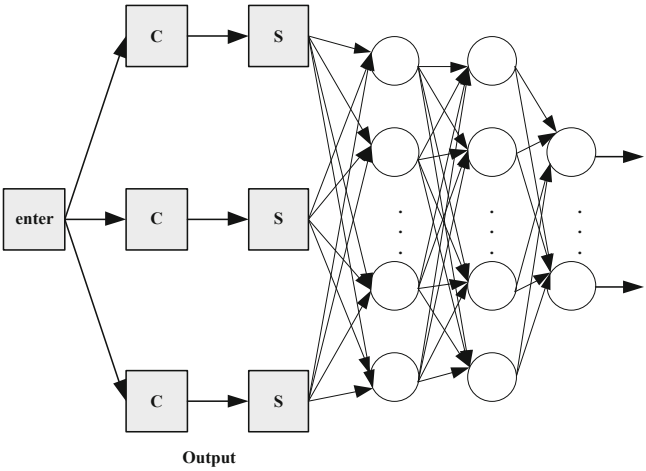

Fig. 1. Structure of porous convolution neural network

The network structure includes an input layer, an output layer and a deconvolution subnet [10–12]. In a convolutional network, the hidden layer of the network is composed of feature maps. The input layer of the network does not limit the image size, so there is no need to process the image size. An image containing noise can be input from the input layer.

After the image is input into the porous convolution neural network, the convolution check image of (5×5) is used for convolution operation, adding bias for the image to

be sampled to form CI layer. After adding the offset, the network contains 32-bit feature map, then the number of parameters to be trained is $5 \times 5 \times 1 \times 32 + 32$, and then the convolution kernel is reduced to (3×3). On this basis, the number of feature graphs is increased, and the feature graphs are connected with the feature graphs of the upper layer to form a C2 layer with 64 feature graphs, and 64 feature graphs are generated correspondingly. The number of parameters to be trained is $1 \times 1 \times 64 \times 32 + 32$. On the basis of C2 layer generated by deconvolution and operation of CI layer, the size of convolution kernel is increased to (5×5), and the size of output feature map is set to 1. At this time, the output of deconvolution is the denoised image output from network output layer.

Based on the above-mentioned processing of image spatial structure information and noise, the feature fusion of image spatial structure depth information is defined according to the above-obtained image texture features, color features, and shape feature properties, as shown in the following figure:

In image fusion, starting from point $(0, 0)$, the square of (3×3) is used to move from top to bottom and from left to right. If the structural element is the same as one of a, B, C, D and E, the value is retained. If it is different, it is abandoned and continues to fuse.

On the basis of the above image fusion, the similarity measurement of image spatial structure information is carried out. A feature value is extracted from each image in the data set, which is recorded as $K = \{k_1, k_2, \cdots k_n\}$ and stored in the porous convolution neural network. The deeper characteristic value of the image to be represented is extracted and recorded as $W = \{w_1, w_2, \cdots w_n\}$. the absolute distance between each point in the image is measured by Euclidean distance. The calculation formula is as follows:

$$D(d, g) = \sqrt{\sum_{i=1} h}(d - t) \tag{12}$$

In formula (12), $D(d, g)$ represents the absolute distance between image points in the multidimensional space, $\sqrt{\sum_{i=1} h}$ represents the smaller weight coefficient in the image point value, and $d - t$ represents the image data set. According to the above calculation, the image color, shape, and texture features are fused, the fusion result is represented in a perceptual frame, and the representation of the stored target is activated.

6 Implementation of Feature Extraction of Students' Ideological and Political Learning Behavior

On the basis of the detailed representation of the spatial structure information of the above images, the learner behavior features are extracted, and the temporal action detection block diagram is as follows (Fig. 2):

The purpose of the timing detection subnet is to extract timing segments that may have actions. Here, for the feature map of $512 \times \frac{Long}{8} \times \frac{High}{16} \times \frac{Wide}{16}$ generated by the feature extraction subnet, the anchor frame mechanism is first used. According to the no free lunch theorem, the anchor frames are evenly distributed at time $\frac{L}{8}$ in the domain,

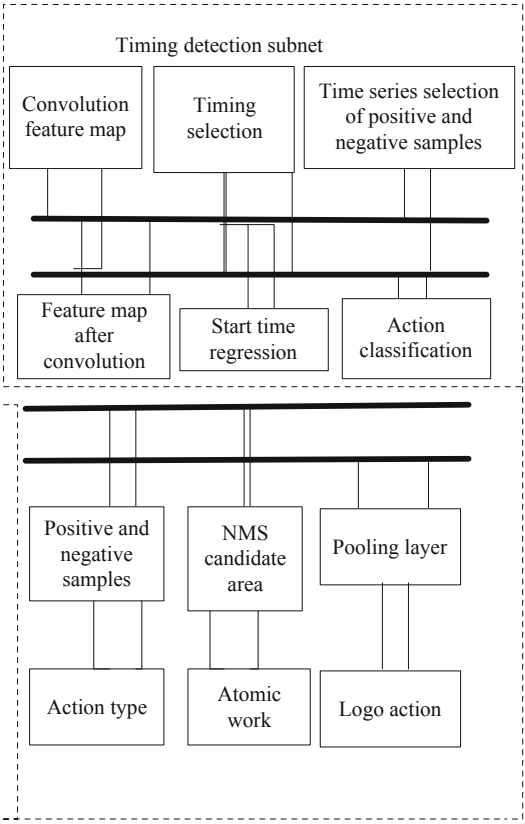

Fig. 2. Overall block diagram of sequential action detection

each anchor box generates k candidate timing points of different sizes. Then, in order to obtain the center position offset and the length of each candidate sequence at each timing point, the spatial feature map of $512 \times \frac{High}{16} \times \frac{Wide}{16}$ is passed through a $3 \times 3 \times$ convolution kernel and a 3D pooling layer to 1×1. Finally output $512 \times \frac{long}{8} \times 1 \times 1$. For the candidate timing to determine whether it is a positive sample or a negative sample, use IOU to calculate when determining. IOU is the number of overlaps between the candidate timing frame and the label. When IOU > 0.7, it is determined as a positive sample; when IOU < 0.3, Determined as a negative sample.

On this basis, the action classification, the classification process is shown in the following figure (Fig. 3):

Feature extraction is an important process of human body target detection. The extraction is represented by vector value and function value. The more vector value and function value describing the characteristics of the target, the greater the amount of information, the richer the details and key information, the more sufficient the detected target is. At the same time, the data dimension will be increased and the calculation difficulty will be increased. In the feature extraction, considering that different scenes

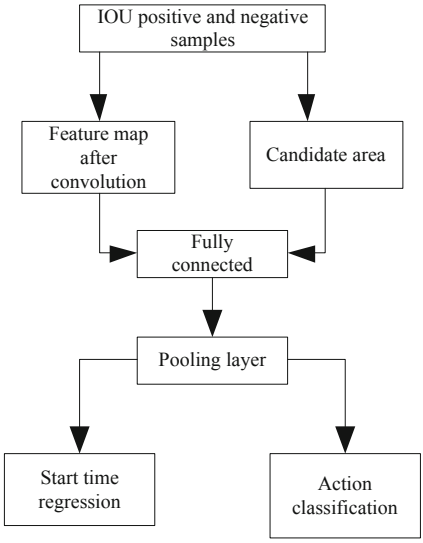

Fig. 3. Action classification subnet

extract different surface features, the extracted features are classified to better and faster targets. For feature description, avoid big data and high-dimensional data, and make the calculation simple and fast. In human object detection, in order to make the results intuitive and more in line with human visual characteristics, the features are divided into two categories, intuitive features and abstract features, as shown in the following figure:

Table 1. Feature classification

Characteristic category	Characteristic subclass	Examples of features
Intuitive features	Pixel features	Color and morphological characteristics
	Attitude features	Distance, angle, manikin
	Action characteristics	Speed, direction, trajectory
Abstract features	Projection features	POI MNU, etc.
	Nonprojective features	Mcvb, JDH, etc.
	Quantitative characteristics of transformation	Wavenumber transformation

According to Table 1, it can be seen that the features are divided into two categories: intuitive features and abstract features. There are some sub class features between them. Among them, the extraction of projection features is to find a linear transformation according to the target, which can lower the data value and complete the feature extraction.

The process of feature extraction plays an important role in the human object measurement. After the processing of the target, the statistical classification of the target behavior, the human behavior and the interaction between the scene are recognized, and the human target detection is completed.

The trajectory of the target can be regarded as a directed curve. The trajectory analysis method is used to reflect the motion path of the moving human body. Firstly, the data set is used to represent the trajectory direction through the string code. In the l-string code, the curve points are represented by the direction coordinate points, and the points that are not in the coordinates are treated with the approximate points, which are used as the data of the motion trajectory

$$e = \left[(x_1, y_2), (x_2, y_2) \cdots, \mathrm{x_n}, \mathrm{y_n}\right] \tag{13}$$

Among them, $\mathrm{x_n}$, $\mathrm{y_n}$ represents the coordinate position of the moving target in the nth frame of the image, and e represents the entire motion trajectory. The obtained motion trajectory is analyzed and processed, and the formula is as follows:

$$\lambda = (x_m - y_{m-1})^2, (x_n - y_{n-1})^2 \tag{14}$$

On the basis of obtaining the motion trajectory, the corner distance is calculated, assuming that the coordinate of the motion trajectory is $x_m - y_{m-1}$, the distance is $(x_n - y_{n-1})$, and λ is the distance of the target point. This calculation does not do orientation analysis.

In the process of motion trajectory analysis, the analysis of target point distance is an important feature in trajectory analysis. The premise of understanding human behavior is to detect and obtain human body trajectory.

Finally, HRG algorithm is used to recognize the target through calculation, and the human behavior is understood mainly according to the relevant gradient and edge position in the image. Firstly, the image gradient is calculated, and pixels are divided by gradient calculation. The gradient calculation formula is as follows:

The main process is as follows (Fig. 4):

The algorithm process is shown in the figure. When the image is detected, the gradient direction of each pixel and the projection cell are calculated, the window is evenly divided into equal intervals, the adjacent pixel values are counted, the cells are normalized, and the HRG features are collected, And finally complete the test. On this basis, the above process realizes the extraction of students' ideological and political learning behavior characteristics.

7 Experimental Comparison

In order to verify the effectiveness of the feature extraction method of students' Ideological and political learning behavior based on convolutional neural network, the experimental comparison is carried out, and the method is compared with the traditional method, and the accuracy and extraction time of the two methods are compared. The comparison results are as follows.

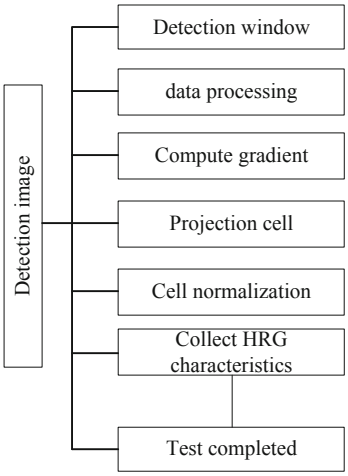

Fig. 4. HRG algorithm process diagram

7.1 Accuracy Comparison of Feature Extraction

The comparison results between the traditional method and the feature extraction method in this study are shown in the following figure (Fig. 5):

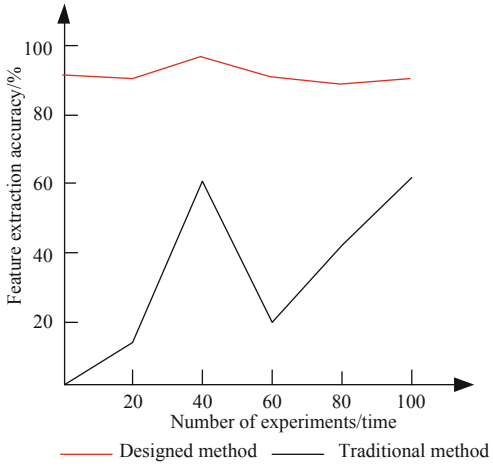

Fig. 5. Comparison of the accuracy of behavior feature extraction

It can be found in the figure that the traditional feature extraction method is extracted by only 60%, the extraction accuracy is low, and the method of extracting the accuracy rate can be up to 95%, while the traditional method is less accurate in several experiments. Students 'thinking of students' thinking behavior based on convolutional neural networks.

7.2 Comparison of Feature Extraction Time

This paper analyzes the feature extraction time between the method and the traditional method, and the results are as follows (Fig. 6):

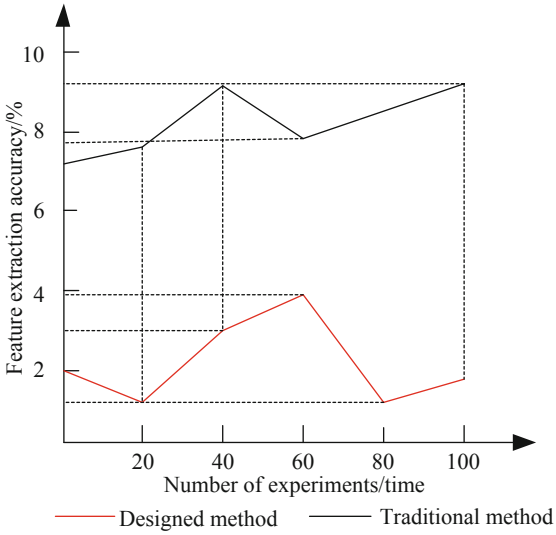

Fig. 6. Feature extraction time comparison

Through the figure above, the extraction time of the reconciled neural network based on convolutional neural network, the extraction time of the convolutional neural network, in 4S, the extraction time of the traditional method is within 9.5 s, based on the convolutional neural network, student thinking The extraction time of the learning behavior feature extraction method is significantly less than the extraction time of the traditional extraction method, which proves the effectiveness of the characteristic extraction method of the study.

8 Conclusion

This paper designs a feature extraction method of students' ideological and political learning behavior based on convolutional neural network, and verifies the effectiveness of this research method through experiments. The reason for the better results of the method designed this time is that the quality of action extraction has a great influence on the positioning effect of subsequent actions in the sequential action detection. Improving the quality of the action extraction in the video helps to improve the overall efficiency of action timing recognition. However, due to the limitation of research time, the method of this research still has certain shortcomings. In the follow-up research, the method of this research needs to be further optimized.

References

1. Feng, J., Cai, S., Ma, X.: Enhanced sentiment labeling and implicit aspect identification by integration of deep convolution neural network and sequential algorithm. Clust. Comput. **22**(6), 1–19 (2019)
2. Wang, X., Zhong, X., Xia, M., et al.: Automatic carotid artery detection using attention layer region-based convolution neural network. Int. J. Humanoid Rob. **16**(4), 1097–1105 (2019)
3. Zhang, F., Cai, N., Wu, J., et al.: Image denoising method based on a deep convolution neural network. IET Image Proc. **12**(4), 485–493 (2018)
4. Liu, C., Zhang, X., Hu, Q.: Image super resolution convolution neural network acceleration algorithm. Guofang Keji Daxue Xuebao/J. Natl. Univ. Defense Technol. **41**(2), 91–97 (2019)
5. Timilsina, S., Aryal, J., Kirkpatrick, J.B.: Urban tree cover changes using object-based convolution neural network (OB-CNN). Remote Sens. **12**(18), 1–27 (2020)
6. Liu, S., Lu, M., Li, H., et al.: Prediction of gene expression patterns with generalized linear regression model. Front. Genet. **10**, 120 (2019)
7. Liu, S., Li, Z., Zhang, Y., et al.: Introduction of key problems in long-distance learning and training. Mob. Netw. Appl. **24**(1), 1–4 (2019)
8. Liu, S., Glowatz, M., Zappatore, M., et al. (eds.): e-Learning, e-Education, and Online Training, pp. 1–374. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93719-9
9. Tu, S., Rehman, S.U., Waqas, M., et al.: Optimisation-based training of evolutionary convolution neural network for visual classification applications. IET Comput. Vis. **14**(5), 259–267 (2020)
10. Liu, Z., Shi, S., Duan, Q., et al.: Salient object detection for RGB-D image by single stream recurrent convolution neural network. Neurocomputing **363**(21), 46–57 (2019)
11. Wang, K., Hinz, J., Zhang, Y., et al.: Parallel channels for motion feature extraction in the pretectum and tectum of Larval Zebrafish. Cell Rep. **30**(2), 442–453 (2020)
12. Rankin, J., Rinzel, J.: Computational models of auditory perception from feature extraction to stream segregation and behavior. Curr. Opin. Neurobiol. **58**, 46–53 (2019)