

Research Article

Online Classroom Teaching Quality Evaluation System Based on Facial Feature Recognition

Fang Yuan^{1,2} and Yong Nie ¹

¹School of Education, Shaanxi Normal University, Xi'an, Shaanxi 710062, China

²School of Education, Xi'an International Studies University, Xi'an, Shaanxi 710128, China

Correspondence should be addressed to Yong Nie; nieyong@snnu.edu.cn

Received 8 November 2021; Revised 21 November 2021; Accepted 29 November 2021; Published 21 December 2021

Academic Editor: Tongguang Ni

Copyright © 2021 Fang Yuan and Yong Nie. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of computer big data technology, online education in the form of online courses is increasingly becoming an important means of education. In order to objectively evaluate the teaching quality of online classroom, a teaching quality evaluation system based on facial feature recognition is proposed. The improved (MTCNN) multitask convolutional neural network is used to determine the face region, and then the eye and mouth regions are located according to the facial proportion relationship of the face. The light AlexNet classification based on Ghost module was used to detect the open and close state of eyes and mouth and combined with PERCLOS (percentage of eye closure) index values to achieve fatigue detection. Large range pose estimation from pitch, yaw, and roll angles can be achieved by easily locating facial feature angles. Finally, the fuzzy comprehensive evaluation method is used to evaluate students' learning concentration. The simulation experiments are conducted, and the results show that the proposed system can objectively evaluate the teaching quality of online courses according to students' facial feature recognition.

1. Introduction

Online education, as a brand new education model, has been gradually welcomed by students and parents due to its characteristics of openness and diversity [1]. Online education market has also ushered in a vigorous development, while there are still numerous shortcomings in the actual use process. The biggest deficiency lies in the poor supervision effect of teachers on students in the classroom [2]. In the traditional teaching process, teachers and students are in a limited space. The position of the podium is very good for quick observation of the whole classroom. In the course of class, teachers can control the whole room and effectively supervise students with just a glance. But in online education, the limits of space are broken down. Most of the teacher's attention is paid to the lecture; thus, the rest of the attention cannot effectively control the state of all students, let alone play a supervisory role. For students, online education offers many choices and flexible hours, while the

degree of self-discipline is very high. The analysis results of the statistical data of some online education platforms show that there is an obvious phenomenon of learners' poor autonomy in online classroom. This results in a high enrolment rate but a low course pass rate for all kinds of courses [3, 4]. During the COVID-19 pandemic in 2020, some parents reported that the learning effect of their children's online classes at home was very low. This not only causes many students to waste valuable learning time, but also undermines teachers' control over teaching progress and quality.

The COVID-19 pandemic is the first large-scale use of online education to replace traditional education in a short period of time, though it cannot completely replace traditional education at present [5]. However, with the emergence and application of 5G technology, online education has gradually become an important development trend that complements traditional education. Therefore, considering that online education teachers cannot effectively supervise

students, the exploration of online classroom teaching quality evaluation system has become an important task of online education development.

Online education has the characteristics of loose structure and open distance teaching environment, in which online learning effect evaluation is an important part. It is very difficult to objectively evaluate the teaching quality of the course via these characteristics [6]. Literature [7] discusses how to improve the learning effect and efficiency of learners through positive emotion and positive emotion in online education. Literature [8] analyses how teachers channel and dissolve students' negative emotions through positive emotions, so as to improve students' learning enthusiasm. Literature [9] builds an emotional interaction model of social learning network, which can identify learners' emotional states in online education. Emotional states include six categories: pleasure, pain, tension, calm, surprise, and disgust. Literature [10] designs a teaching feedback strategy construction method based on emotion learning ontology. In the current research environment, quantitative analysis of student behaviour is an important link of online education. Based on the hybrid teaching mode combining online and offline, paper [11] established online evaluation and offline evaluation systems. The evaluation system focuses on the learning process and teachers can take intervention measures according to the early warning situation.

This paper presents an online classroom teaching quality evaluation system based on facial feature recognition. The innovations and contributions of this paper are listed below.

- (1) Face regions are determined by an improved multitask convolutional neural network, and then eye and mouth regions are located according to the facial proportion relationship of the face. The light AlexNet classification based on Ghost module was used to detect the open and close state of eyes and mouth and combined with PERCLOS index values to achieve fatigue detection.
- (2) Complete the face pose estimation from pitch, yaw, and roll angles in a wide range through easily located facial feature angles.
- (3) Finally, the fuzzy comprehensive evaluation method is used to quantitatively evaluate students' learning concentration.

The simulation results show that the system can effectively evaluate students' concentration in class according to the results of face detection, so as to evaluate the teaching quality of online education.

The chapter structure of this paper is as follows. Section 2 focuses on the proposed algorithm model in this paper. Section 3 is experiment and analysis. Section 4 is the conclusion.

2. The Proposed Model in This Paper

2.1. Fatigue Detection Model. Fatigue detection based on human facial features usually needs to determine the location of the face and eye-mouth and detect the opening and

closing state of the eye-mouth. In order to achieve the above functions, this paper uses the improved MTCNN (Multitask Cascaded Convolutional Networks) to complete face detection and key point positioning. Then, the improved AlexNet was used to identify the open and closed state of the eyes and mouth. Finally, the fatigue discrimination was performed according to PERCLOS and PMOT parameters. The fatigue detection process is shown in Figure 1.

2.1.1. Key Point Positioning Based on Improved MTCNN.

MTCNN is a face detection and alignment algorithm based on deep learning, which uses image pyramid to detect faces at various scales [12]. Different from multitarget face detection, fatigue detection only needs to precisely locate the human face region. Therefore, the image pyramid part of MTCNN network is improved to achieve face detection and key point location quickly and accurately. Then, after locating the key points, the eye and mouth regions are obtained by using the proportion relation of three courtyards and five eyes of human face. Three courtyards here refer to the length of the face proportion, and the length of the face is divided into three equal points. Five eyes here refer to the proportion of the width of the face, and the width of the face is divided into five equal parts with the length of the eye as the unit.

MTCNN is made up of Proposal Network (P-Net), Refine Network (R-Net), and Output Network (O-Net). The idea of candidate frame plus classifier is used to detect face quickly and efficiently. In order to detect faces of different scales, MTCNN reduces the original image to different sizes to generate image pyramids. The original image is compressed to a certain scale, and the image is traversed with a candidate box of fixed size 12 pixels by 12 pixels. Loop until the reduced image is smaller or wider than the side length of the candidate box. At this point, a picture with a size of 12 pixels \times 12 pixels and channel number of 3 is obtained as the input of P-Net. P-Net uses a convolution kernel with a size of 3×3 to extract image features through convolution operation. Maximum pooling operation was used to remove redundancy. Face classification is used to determine whether the region is a face. Face regions were initially located using bounding box regression and a facial landmark localization. These fields are inserted into R-Net, which will repeat the operations of P-Net. Most of the interference is filtered out; then the reliable face area is retained and input into O-Net. P-Net takes advantage of the feature that the full-convolution output size is 1 pixel \times 1 pixel and the number of channels is 32. R-Net uses a 128-dimensional full connection layer after the last convolutional layer, which preserves more image features. O-Net performs more refined face discrimination, border regression, and key point positioning and finally outputs the coordinates of the face region and 5 feature points. The five characteristic points are left eye midpoint, right eye midpoint, nose tip, left corner of mouth, and right corner of mouth.

The image pyramid constructed by MTCNN contains a large number of images. Sending it into the network to detect all the face areas in the image would take a lot of time. In the

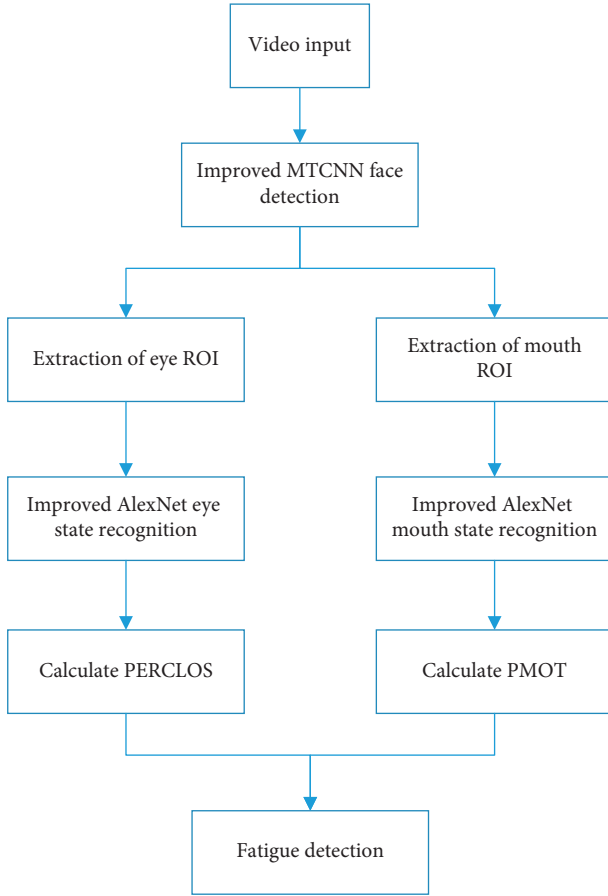


FIGURE 1: Fatigue detection process.

fatigue detection scene, it is necessary to eliminate the interference of redundant faces and accurately locate human faces. Considering that the human face region occupies a large proportion of pixels in the video frame, this paper amplifies the minimum face size in MTCNN network according to the proportion of human face in the image. The fixed reduction scale is reduced and the first reduction and traversal are skipped. This allows the improved MTCNN network to accurately locate human facial areas. At the same time, it can filter out irrelevant faces in the background and greatly reduce the face detection time.

It is assumed that the coordinate of the left eye obtained by MTCNN is $t_1(\alpha_1, \beta_1)$ and the coordinate of the right eye is $t_2(\alpha_2, \beta_2)$. With extraction of human eye region according to face proportion relationship, its corresponding relationship is shown as

$$\begin{cases} f_1 = \sqrt{(\alpha_1 - \alpha_2)^2 + (\beta_1 - \beta_2)^2}, \\ \omega_1 = 0.5f_1, \\ \varphi_1 = 0.5f_1, \end{cases} \quad (1)$$

where f_1 is the distance between the eyes. ω_1 is the width of the eye area. φ_1 is the height of the eye area.

It is assumed that the coordinates of left and right corners of human mouth obtained by MTCNN are $m_1(\alpha_1, \beta_1)$ and $m_2(\alpha_2, \beta_2)$, respectively. The mouth region of

human is extracted according to the proportion relation of face, and its corresponding relation is

$$\begin{cases} f_2 = \sqrt{(\alpha_1 - \alpha_2)^2 + (\beta_1 - \beta_2)^2}, \\ \omega_2 = f_2, \\ \varphi_2 = f_2, \end{cases} \quad (2)$$

where f_2 is the distance between the left and right corners of the mouth. ω_2 is the width of the region of interest of the mouth. φ_2 is the height of the mouth area.

Key points were located for the examples in NTHU-DDD dataset [13], and the regions of interest of eyes and mouth were obtained according to the proportion of three courts and five eyes, as shown in Figure 2.

2.1.2. Eye and Mouth State Recognition. The method of eye and mouth state recognition based on manual feature extraction is affected by the shooting angle, shooting distance, and individual differences and has poor robustness. In this paper, the improved AlexNet is used to recognize the open and closed state of the eyes and mouth, which avoids the complex preprocessing operation of the image and has strong robustness.

Ghost module [14] is a lightweight neural network unit. In order to make the fatigue detection model obtain better real-time performance at the edge end, Ghost module is used in this paper to replace all convolution operations in AlexNet. Due to the redundancy of feature images output by deep convolutional neural network, these similar feature images make the feature extraction ability of convolutional neural network stronger. Therefore, Ghost module uses a simple linear operation to obtain more similar feature graphs to improve CNN performance. Ghost module uses a small amount of conventional convolution to get eigen features and then the obtained eigen features through depthwise convolution such a simple linear operation to generate Ghost features. Finally, the intrinsic feature graph and Ghost feature graph are spliced to get the final output. Compared with the direct use of conventional convolution, Ghost module, in the guarantee of accuracy at the same time, greatly reduces the amount of calculation.

AlexNet [15] showed excellent results in image classification based on 8-layer network structure. AlexNet is improved in this paper to classify the open and closed state of eyes and mouth. Since the image of eyes and mouth occupies fewer pixels, the input size of AlexNet is compressed from 224 pixels \times 224 pixels and channel number 3 to 24 pixels \times 24 pixels and channel number 3. Modify 11 \times 11 and 5 \times 5 convolution kernels to 3 \times 3 convolution kernels. The stride of maximum pooling operation was optimized to avoid too small size of feature graph. AlexNet is only used for eye-mouth state 4 classification, and only the first full connection layer in AlexNet is retained in the model. The output dimension of the full connection layer is changed from 2 048 to 128. Finally, Softmax regression function is used to output the probability that the sample belongs to the open mouth state. The model retains the first layer of traditional convolution to extract image features



FIGURE 2: ROI positioning of eyes and mouth. (a) Normal state. (b) Yawning state. (c) Slow blink with nodding. (d) Sleepy state.

comprehensively. The other convolution operations are replaced by Ghost module, making the network lightweight. The improved AlexNet architecture is shown in Figure 3.

2.1.3. Fatigue Detection. When a person is in a state of fatigue, there will be a series of physiological reactions, such as long closed eyes, yawning and so on. By calculating PERCLOS and PMOT parameters of continuous video frames, the improved AlexNet model can judge the fatigue state of human with threshold value.

PERCLOS is a physical quantity proposed by Carnegie Mellon Institute to measure fatigue. The correlation analysis of 9 fatigue parameters, including PERCLOS parameters, blinking frequency, and yawning parameters, shows that the correlation between PERCLOS parameters and fatigue state is the highest [16]. PERCLOS parameter represents the percentage of eye closure time in unit time, and the calculation equation is shown as follows:

$$P_{\text{PERCLOS}} = \frac{\sum_x^M h_x}{M} \times 100\%, \quad (3)$$

where M is the total number of video frames per unit time. h_x is the number of closed frames. $\sum_x^M h_x$ is the total number of frames closed per unit time. In the normal state, the value of the PERCLOS parameter is small. When people are in

fatigue state, the number of closed eye frames increases, and the value of PERCLOS parameter increases.

The PMOT parameter is similar to the PERCLOS parameter and represents the percentage of open mouth time per unit time. Its calculation equation is as follows:

$$P_{\text{PMOT}} = \frac{\sum_x^M h_x}{M} \times 100\%, \quad (4)$$

where M is the total number of video frames per unit time. h_x is the number of open mouth frames. $\sum_x^M h_x$ is the total number of open mouth frames per unit time. In normal process, the value of human PMOT parameter is small. When a person yawns, the number of open mouth frames increases and the value of PMOT increases.

2.2. Face Pose Estimation Algorithm. Face deflection axis is shown in Figure 4. Among them, the deflection around the X axis is called pitch, the deflection around the Y axis is called yaw, and the deflection around the Z axis is called roll.

In this paper, Adrian Bulat face feature point locator is used to complete face feature point positioning [17], as shown in Figure 5. The positioner is suitable for feature location of rotating face in plane. In addition to the visible feature points, those blocked or invisible face feature points can also be located.

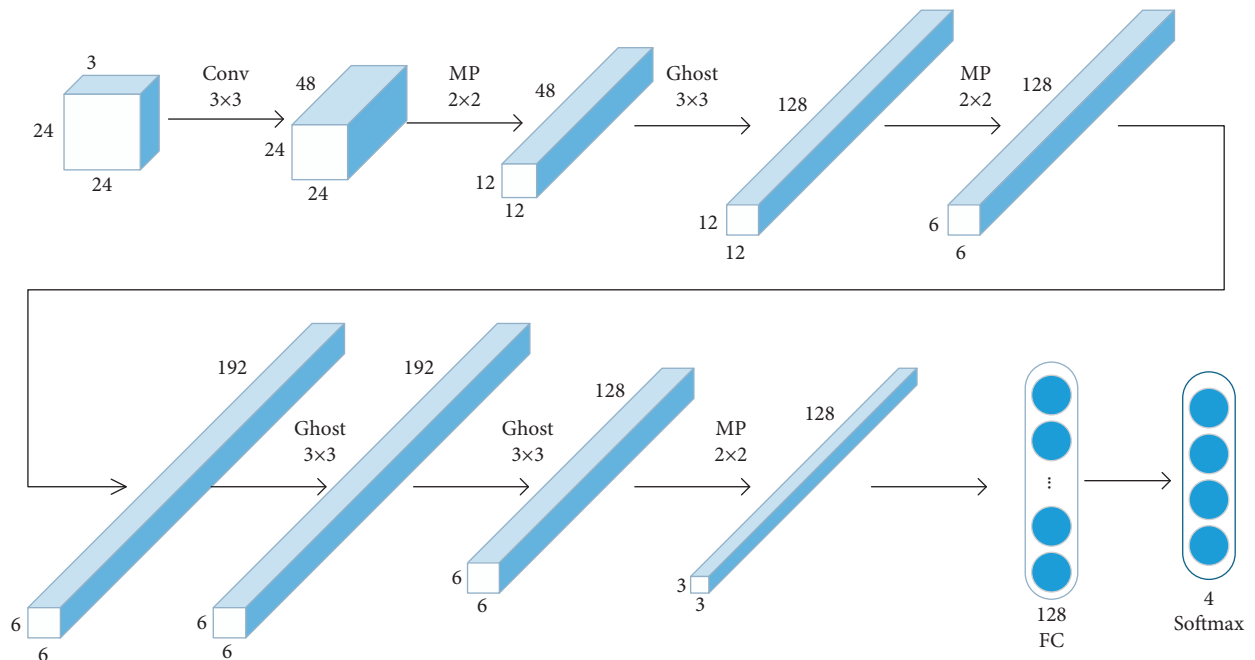


FIGURE 3: Improved AlexNet architecture.

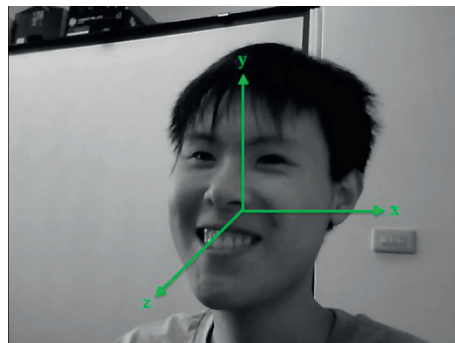


FIGURE 4: Face posture and its coordinate axes.

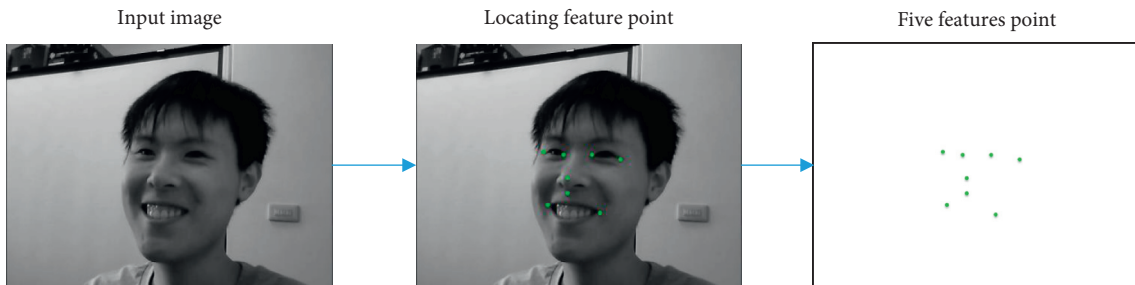


FIGURE 5: The process of obtaining facial feature points.

Based on facial features, the face pose estimation process adopted in this paper is shown in Figure 6. In order to reduce the estimated parameters in the loss function, the algorithm transforms the process of estimating face rotation around three axes into the process of searching the best rotation angle around X and Y axes within a certain rotation range around Z axis of sparse model. Thus, the roll angle parameter in the loss function is eliminated.

By aligning the subnasal point on the 3D model with the subnasal point on the image, the constraint model can only rotate around the subnasal point to eliminate the translation parameter in the loss function. This makes the loss function retain only the scaling factor, pitch angle, and yaw angle.

Let s be the global dimension parameter of the 3D model; w_α and w_β are the translation parameters in the X and Y directions of the 3D model after the parallel projection to the

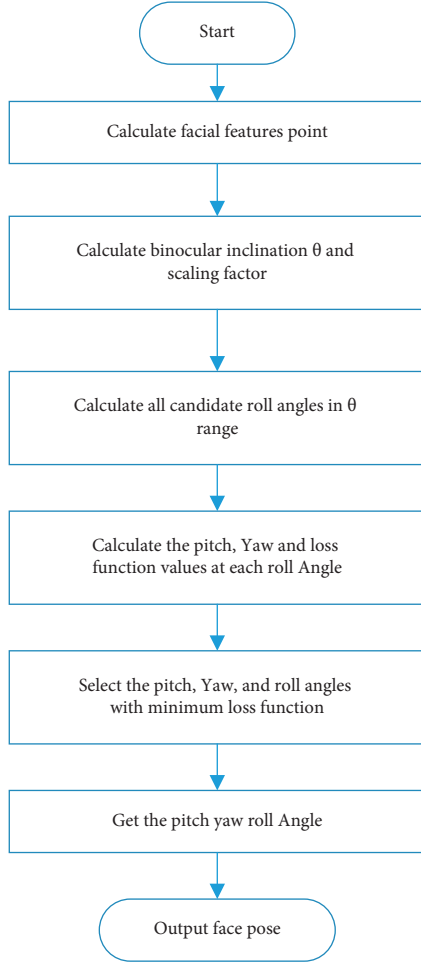


FIGURE 6: The process of face pose estimation.

XY plane. If face roll angle γ is known, the following method is used to estimate face depth direction deflection angles α

$$\min H(s, \alpha, \beta) = \min \left(\sum_{x=1}^t \left\| P_{3D} R_{3D} S_{3D} (v_x^{3D} - v_u^{3D}) + (v_u^{2D} - v_x^{2D}) \right\|^2 \right) + \frac{r_z}{s}. \quad (9)$$

$r_z > 0$ is the barrier factor. The modified Newton method [18] is used to calculate the deflection parameters α and β of pitch and yaw of the image face satisfying equation (9) at the specified γ angle, as well as the scaling coefficient s of 3D face model v .

In this paper, the estimations of face in-plane rotation angle and out-plane deflection angle were combined, and the final estimation of face deflection angle around each coordinate axis was obtained by searching for the best deflection angles α and β are within the range of $\theta \pm 90^\circ$ based on the angle θ of the two-eye center line.

2.3. Fuzzy Comprehensive Evaluation. In practical problems, objective decision is often a comprehensive decision of many

and β . The subnasal points of 3D model v were recombined and fixed with the subnasal points of human face on the image. Then, s , α , and β were adjusted to align other feature points on the image with corresponding points on 3D model v after 2D projection (meeting the minimum sum of squares of distance). The equation for the sum of the squares of the minimum distance is as follows:

$$\min h(s, \alpha, \beta) = \min \left(\sum_{x=1}^t P_{3D} R_{3D} S_{3D} v_x^{3D} + W_{2D} - v_x^{2D} \right), \quad (5)$$

$$\begin{cases} W_{2D} = v_u^{2D} - P_{3D} R_{3D} S_{3D} v_u^{3D}, \\ s > 0, \\ \alpha, \beta, s \in R, \end{cases} \quad (6)$$

where t is the number of alignment points, and v_x^{2D} is the alignment point on the face image.

$$v_x^{2D} = (\alpha_x^{2d}, \beta_x^{2d}, 0)^T, \quad (7)$$

v_x^{3D} is the corresponding alignment point on model v .

$$v_x^{3D} = (\alpha_x^{3d}, \beta_x^{3d}, \gamma_x^{3d})^T. \quad (8)$$

v_u^{2D} and v_u^{3D} are the subnasal points on the face image and v , respectively. W_{2D} is the translation vector after the projection of v . P_{3D} is v orthographic matrix. R_{3D} is v rotation matrix. S_{3D} is v global scaling matrix.

Substitute equation (6) into equation (5), and construct the augmented objective function (loss function) using the interior point penalty function method; then it can be obtained as follows:

factors. Some attributes are ambiguous, cannot be quantified, and cannot be judged simply by "good" or "bad." Fuzzy comprehensive evaluation decision is a comprehensive evaluation method based on fuzzy mathematics [19]. The basic idea is to use fuzzy linear transformation principle and maximum membership principle. Fuzzy attributes are quantified by membership function and evaluated by traditional quantitative evaluation method. In this way, the quantitative and qualitative factors in the problem can be uniformly dealt with, and the differences between fuzzy attributes can be taken into account.

The evaluation objective of this paper is learning concentration. It cannot be quantified in the academic world and belongs to the fuzzy goal. Facial orientation and fatigue degree are also fuzzy in evaluation, so fuzzy comprehensive

evaluation is adopted in this paper. The mean value of horizontal facial deflection angle, vertical facial deflection angle, times of eye closure, and times of yawn per unit time are the first level factors. Facial orientation and fatigue were the second factors. Therefore, this paper adopts fuzzy comprehensive evaluation oriented to multilayer factor evaluation, that is, multilevel fuzzy comprehensive evaluation.

2.3.1. Factor Evaluation of the First Layer. The head posture score was divided into turning head score and raising head score, and the minimum value was taken for the comprehensive score. Firstly, the mean of left (right) head turn angle θ_1 and the mean of raised (low) head angle θ_2 were calculated. According to the left and right rotation range of human cervical vertebra joint is $\pm(60^\circ \sim 80^\circ)$. 70° is taken as the maximum left (right) head rotation angle in this paper. The smaller the angle θ_1 is, the higher the score is.

$$z_1 = \begin{cases} 1 - \frac{|\theta_1|}{70^\circ}, & 0^\circ \leq |\theta_1| \leq 70^\circ, \\ 0, & |\theta_1| > 70^\circ. \end{cases} \quad (10)$$

The angle of raising (lowering) head of human cervical vertebra joint is $-60^\circ \sim +60^\circ$. When the angle is greater than 0° , it is judged as the head, and when the angle is less than 0° , it is judged as the head. The closer the angle θ_2 is to 0, the closer the head is to the ground, indicating that the learner is looking straight ahead. So the closer the angle θ_2 gets to zero, the higher the score is.

$$z_2 = \begin{cases} 1 - \frac{|\theta_2|}{60^\circ}, & 0^\circ \leq |\theta_2| \leq 60^\circ, \\ 0, & |\theta_2| > 60^\circ. \end{cases} \quad (11)$$

The comprehensive score takes the minimum value as

$$\rho_1 = \min(z_1, z_2). \quad (12)$$

In this paper, a comprehensive score of PERCLOS value [20], average length of eye closure, and yawning frequency was used to evaluate fatigue. The comprehensive evaluation of concentration included τ times measurement of eye closure, yawn, and head rotation. The total time of τ times detection is W . PERCLOS value t_1 was calculated according to the number of eye closure T_1 and detection τ .

$$t_1 = \frac{w_1}{\tau}. \quad (13)$$

The average closed eye duration (t_2) was calculated from the total time (W) of τ times detection.

$$t_2 = \frac{W}{w_1}. \quad (14)$$

Yawning frequency t_3 was calculated according to the number of yawns w_2 .

$$t_3 = \frac{w_2}{\tau}. \quad (15)$$

The weights of the three parameters are determined according to their importance. Take the weights of the three as 1, 0.8, and 0.5, respectively, and the comprehensive fatigue score ρ_2 can be calculated as

$$\rho_2 = t_1 + 0.8t_2 + 0.5t_3. \quad (16)$$

2.3.2. Second Factor Evaluation. The multilevel comprehensive evaluation model of fuzzy comprehensive evaluation model was used to calculate the comprehensive score of concentration. The scores for the first tier of factors have been calculated. Facial orientation score and fatigue score were grouped into evaluation factors $p = [\rho_1, \rho_2]$. Analytic hierarchy process was used to calculate the weight of each factor.

The eigenvector is $\Omega = [0.653, 0.347]^T$. The comprehensive score of concentration is the sum of the scores of the two factors and the weight product, and the equation is as follows:

$$\rho = P \cdot \Omega. \quad (17)$$

3. Experiment

3.1. System Development. The computer development system of this system is Win10. The programming language for development is Python. The development tool uses Pycharm. The programming environment is Python 3.8. The system is set to perform a test every one second. Every five times is a detection period.

3.1.1. Image Preprocessing. This system uses OpenCV's own function to open the computer camera and obtain images. If the fetch fails, it is fetched again. To judge the resolution of the successfully obtained image, if its length (width) is greater than 700, the image is reduced to 1/3 of the original image. Grayscale processing of image can reduce the complexity of subsequent calculation. The face feature points of the image are detected using the preloaded DLIB library. If the detection is successful, 68 feature points and ash images are returned. If the detection fails, obtain the image again and perform the above steps again.

3.1.2. Head Posture Assessment. This module firstly obtains face feature points outputted by image acquisition and preprocessing module. The feature points of canthus of both eyes, corners of the mouth, tip of the nose, and subnasal points were selected. Binocular inclination, scaling factor, and all candidate roll angles were calculated. The pitch, yaw, and loss function values at each roll angle were calculated using the modified Newton iteration method. Select the pitch, yaw, and roll angles corresponding to the minimum loss function. Finally, the evaluation value of face pose is output.

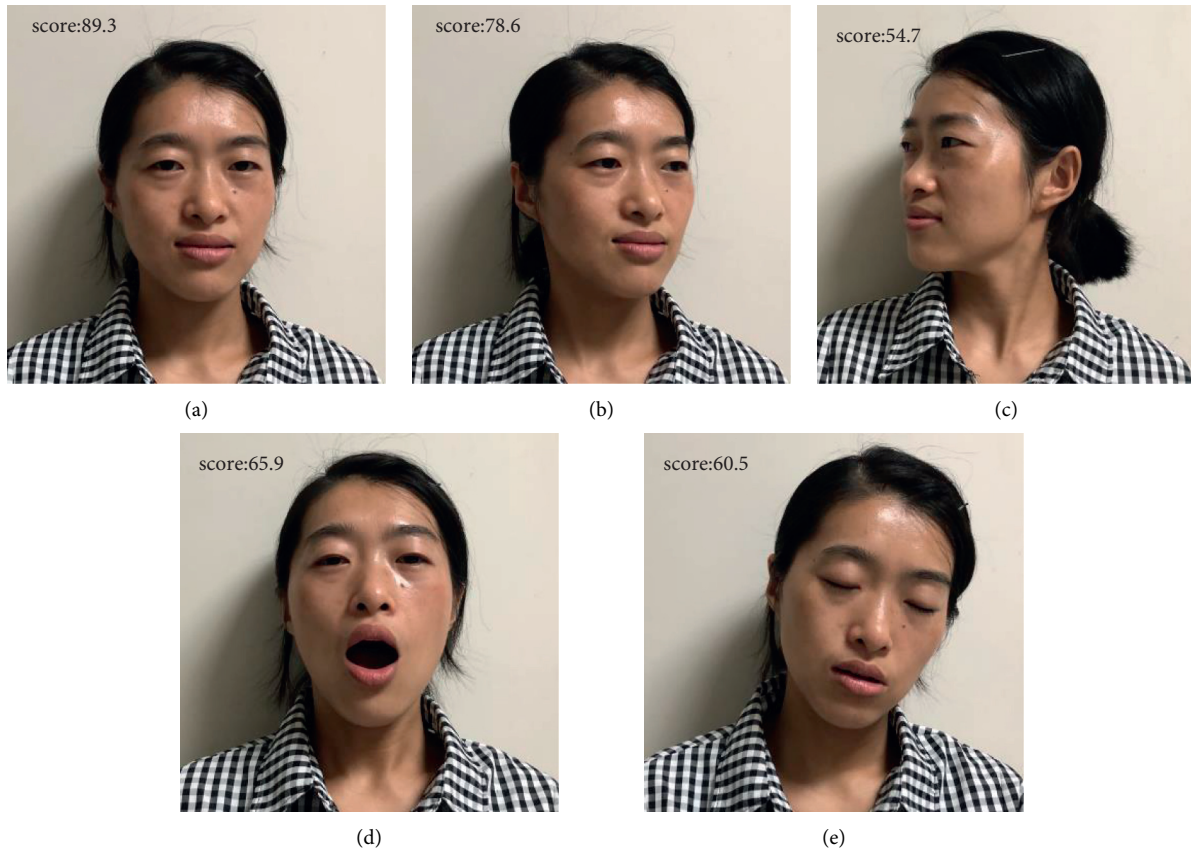


FIGURE 7: Five learning states of learners. (a) Normal state. (b) Mild head deflection. (c) Severe head. (d) Yawning. (e) Being asleep.

3.1.3. Eye Closure Test and Yawn Test. This module firstly obtains image acquisition and preprocessing module outputs face feature points. Select the feature points of left (right) eye and mouth. The mean value of left (right) eye closure was calculated. Check whether the closure is less than or equal to the threshold. If the condition is met, the number of eye closures is increased by 1. The mouth closure is calculated according to the characteristic points of the mouth to determine whether the closure is greater than or equal to the threshold. If the condition is met, the number of yawns is increased by 1.

3.1.4. Fuzzy Comprehensive Evaluation. The fuzzy comprehensive evaluation module firstly determines whether the detection cycle is over. If not, continue monitoring. At the end, the head posture score and fatigue score were, respectively, calculated according to the detected head raising angle, head left-right turning angle, PERCLOS value, average eye closing duration, and yawn frequency. Then the learning concentration score was calculated based on fuzzy comprehensive score.

3.2. System Experiment. In order to objectively evaluate the teaching quality of online classroom, a teaching quality evaluation system based on facial feature recognition is proposed. The effectiveness of the scheme needs to be tested in practical application. Therefore, a detection scheme is

designed in this paper. The subjects were watching a 40-minute teaching video on the computer while the detection system started working. During this process, subjects randomly simulated normal state, nodding state, yawning state, mild head deflection state, and severe head deflection state. It can be seen as Figure 7.

The detection system calculates the mean value of the learning concentration score given by the system to learners in different states. The experimenters were divided into four groups with 20 people in each group. The experimental results are shown in Table 1. For the same learning state simulated by different experimenters, there was no significant difference between the scores given by the system to learners. Compared with the normal state, the scores of learning concentration in other states all decreased to different degrees. In the severe head deflection state and the sleepy state, the decrease of learning concentration score was greater. Combined with the actual situation, severe head deflection and drowsiness will have a very negative impact on learners' concentration. When the head is slightly tilted, it indicates that the learner's attention has shifted to a certain extent, but to a small degree. Therefore, at this time, the system score has decreased relative to the normal state, but the decrease is small. Yawning indicates that the learner's mental state shows fatigue, which will have a great negative impact on their learning concentration. Therefore, the score of the system at this time has a large decrease compared to the normal state.

TABLE 1: System scores of learner groups in different states.

Learner groups	Normal state	Mild head tilt	Severe head deviation	Yawn state	Sleepy state
1	88.2	81.6	56.2	67.2	59.9
2	87.8	81.7	53.5	66.5	52.7
3	93.1	79.3	55.6	57.9	63.3
4	89.9	79.8	53.9	61.4	63.6

It shows that the system can realize the detection of learners' learning concentration to a certain extent through the simulation experiment of the system designed in this paper. The quality of online course teaching can also be evaluated through this system. At the same time, the teacher can obtain students' online learning concentration in real time at the remote end and master the classroom status. In order to improve the quality of teaching, teachers can adjust the teaching plan to help students learn better.

4. Conclusion

As a new education model, online education is gradually welcomed by students and parents because of its openness and diversity. Aiming to objectively evaluate the teaching quality of online classroom, a teaching quality evaluation system based on facial feature recognition is proposed. Firstly, the improved MTCNN network is used to obtain key points of face. The eye and mouth regions were obtained according to the proportion of three courtyards and five eyes. The improved AlexNet based on Ghost module is used to classify the state of eyes and mouth and make fatigue judgment. Then, the face pose estimation is completed by associating Adrian Bulat face feature point locator. Finally, the effectiveness of the system is verified by simulation experiments. The experimental results show that this system has a good effect on the evaluation of learning concentration and can objectively evaluate the teaching quality of online courses. The analysis and optimization of algorithm efficiency is the content of future work.

Data Availability

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

Ethical Approval

The authors declare that they obtained image data from public datasets or image data collected by their team, and they obtained authorization to use portrait images in publicly published articles.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was supported by the Fundamental Research Funds for the Central Universities (Project name: The Study of Need Supportive Teaching under Ubiquitous Learning,

Project no. 2016TS009) and research project of Youth Foundation of Humanities and Social Sciences, MOE (Ministry of Education in China) (Project name: The Study of the Historical Evolution of College Teaching Paradigms and Its Transformational Paths, Project no. 19YJC880124). Yuan Fang is the member of "One Belt, One Road" countries' education developmental strategy research team, The Youth Innovation Team of Shaanxi Universities.

References

- [1] P. Paudel, "Online education: benefits, challenges and strategies during and after COVID-19 in higher education," *International Journal on Studies in Education*, vol. 3, no. 2, pp. 70–85, 2021.
- [2] J. Jia and J. Zhang, "The analysis of online learning behavior of the students with poor academic performance in mathematics and individual help strategies," in *Proceedings of the International Conference on Blended Learning*, July 2019.
- [3] L. Q. Wu, "A comparative study on the effect of online classroom teaching and traditional classroom teaching," *Education Modernization*, vol. 6, no. 9, pp. 276–277, 2019.
- [4] Y. Tu and Y. J. Zhang, "An empirical study on the supervision model of knowledge construction effect in online classroom," *Education Modernization*, vol. 41, no. 3, pp. 64–73, 2018.
- [5] L. Zhou, S. Wu, F. Li, S. Wu, and M. Zhou, "'School's out, but class's on', the largest online education in the world today: taking China's practical exploration during the COVID-19 epidemic prevention and control as an example," *Best Evidence of Chinese Education*, vol. 4, no. 2, pp. 501–519, 2020.
- [6] D. Porcello and S. Hsi, "Crowdsourcing and curating online education resources," *Science*, vol. 341, no. 6143, pp. 240–241, 2013.
- [7] Y. Deng and J. Mo, "Analysis of the role of positive affective factors in network autonomous learning and its training approach," *China Electric Power Education*, vol. 30, no. 32, pp. 52–53, 2014.
- [8] T. Deng, "The improvement of teachers' positive emotion in teaching and the elimination of poor students' negative emotion in learning," *Journal of Hunan University of Science and Engineering*, vol. 36, no. 7, pp. 146–147, 2015.
- [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*, vol. 5, no. 4, pp. 2506–2517, 2018.
- [10] X. Tao and Q. Niu, "Instructional feedback strategy generation algorithm based on an emotion learning ontology," *Computer Engineering and Science*, vol. 37, no. 2, pp. 320–328, 2015.
- [11] J. Chen, Z. Wang, J. Chen, Z. Chen, and H. Zhen, "Design and research on intelligent teaching system based on deep learning," *Computer Science*, vol. 46, no. 6A, pp. 550–554, 2019.
- [12] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

- [13] Y. Ed-Doughmi, N. Idrissi, and Y. Hbali, "Real-time system for driver fatigue detection based on a recurrent neuronal network," *Journal of Imaging*, vol. 6, no. 3, p. 8, 2020.
- [14] 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.
- [15] 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*, vol. 5, no. 4, pp. 2633–2645, 2018.
- [16] D. K. Srivastava and P. K. Tiwari, "Prevention of road accidents due to drowsiness detection using PERCLOS algorithm," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 24, no. 5, pp. 1429–1437, 2021.
- [17] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks)," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1021–1030, Venice, Italy, October 2017.
- [18] S. Zhang, Y. Liu, X. Li, and G. Bi, "Fast ISAR cross-range scaling using modified Newton method," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 3, pp. 1355–1367, 2017.
- [19] D. Yang and C. M. Mak, "An assessment model of classroom acoustical environment based on fuzzy comprehensive evaluation method," *Applied Acoustics*, vol. 127, pp. 292–296, 2017.
- [20] D. Cardone, C. Filippini, L. Mancini et al., "Driver drowsiness evaluation by means of thermal infrared imaging: preliminary results," in *Proceedings of the Infrared Sensors, Devices, and Applications XI. International Society for Optics and Photonics*, Article ID 118310P, California, CA, USA, August 2021.