Research Article

Yurong Huang* and Guang Yang

A study on the application of multidimensional feature fusion attention mechanism based on sight detection and emotion recognition in online teaching

https://doi.org/10.1515/jisys-2023-0096
received July 18, 2023; accepted January 24, 2024

Abstract: Online teaching is not limited by time, but the problem of low learning efficiency is common. To address this problem, the study proposes an attention mechanism for multidimensional feature fusion, which first detects faces, uses a supervised gradient descent algorithm for face feature point detection, and improves the least-squares ellipse-fitting algorithm to detect the open/closed state of human eyes. The sight detection method is also improved, and the fuzzy inference method is used to identify students' emotions, and the modules are fused to achieve multidimensional feature fusion attention detection for online teaching. The study found that the average accuracy rate was 84.5% with glasses and 92.0% without glasses. The research method with glasses had an average time consumption of 17 ms, while the method without glasses took 15 ms, indicating higher detection accuracy and faster real-time performance. The improved approach led to higher recognition accuracy and accuracy rate. The detection accuracy of a single feature and the research method was 74.1 and 91.9%, respectively. It shows that the research method helps in the detection of students' attention in online teaching.

Keywords: face features, human eye opening and closing, line of sight detection, emotion recognition, attention detection, online teaching

1 Introduction

In today's era of rapid technological development, computer technology has provided great convenience to people's lives, and the education sector is gradually combining with computer technology to promote the development of online teaching [1]. Unlike traditional teaching, online teaching does not allow teachers to monitor students' emotional state and attention span because they are not able to communicate with them face-to-face, so they cannot effectively monitor them and make adjustments to the content and teaching style [2]. The incorporation of emotion recognition in online teaching can address the problem of emotion deficit and, combined with attention detection, can facilitate the adjustment of teaching strategies, thus facilitating effective learning for students and effective teaching for teachers [3]. To improve the efficiency of online teaching, it is necessary to strengthen the attention detection of students in the online teaching process. This study uses face feature points, human eye opening and closing, line of sight detection, and emotion recognition to achieve multidimensional feature fusion attention detection to facilitate the adaptation of online teaching

* **Corresponding author: Yurong Huang,** Department of Educational Science and Technology, Anshan Normal University, Anshan, 114007, China, e-mail: huang_yr@outlook.com
Guang Yang: Department of Educational Science and Technology, Anshan Normal University, Anshan, 114007, China, e-mail: yangguang@mail.asnc.edu.cn

strategies. The main contribution of this study is to combine gaze detection and emotion recognition to construct an online teaching detection system that integrates attention mechanisms. It has a high detection accuracy and can help detect the attention situation of students in online teaching, which is beneficial for teachers to grasp the learning situation of students and promote teaching adjustments. The study includes four parts: the first part is a literature review, which summarizes the research on emotion recognition by different scholars; the second part is the research methodology, which includes various elements of attention detection; the third part is the analysis of the results, which focuses on the analysis of the results of multidimensional feature fusion attention detection; and the fourth part is the conclusion, which summarizes the relevant results and points out the shortcomings of the study. The contribution of this study is to facilitate the adjustment of teaching strategies through the detection of multidimensional feature fusion attention.

2 Related works

As computer networks develop, online teaching is getting better and better. However, the current online teaching has low learning efficiency, to identify its learning emotional state to determine the learning situation of students. Scholars in many fields have carried out a lot of research and have achieved corresponding research results.

Cui et al. designed a machine learning-based model for identifying students' emotions in business English courses to address the problem that traditional English teaching models do not focus on students' learning emotions. The results showed that the model facilitated the maintenance of students' learning initiatives and enhanced their motivation to learn [4]. Huang et al. proposed a face recognition method based on students' classroom videos to address the problem of low accuracy of current video-based techniques for analyzing students' learning behavioral states in the classroom. The research results showed that the method was able to identify students' attention and emotional states in class [5]. Wang designed an online learning behavior analysis method by image emotion recognition to address the shortcomings in learning behavior monitoring and teaching effectiveness evaluation in online learning platforms. Results showed that it improved the accuracy of online learning behavior analysis and facilitated the evaluation of teaching effectiveness [6]. Wu and Chen designed a convolutional neural network-based English vocabulary detection model to address the problem of low efficiency of vocabulary recognition in online English vocabulary teaching. Results showed that the model improved the efficiency of English vocabulary recognition and facilitated English vocabulary online teaching [7]. Scholars such as Xuan et al. proposed an ADAM-free IMAM-based IMC macro, which achieves energy-efficient and high-precision MAC computation through brain-inspired computation. The use of time-coded spike neuron circuits instead of analog-to-digital converters enables efficient data conversion and increases computational feasibility [8].

The multidimensional feature fusion attention mechanism based on gaze detection and emotion recognition has been studied differently in different fields, verifying the effectiveness of the method. Tan et al. proposed a short-time emotion recognition system by an impulsive neural network model of spatiotemporal EEG patterns to address the main problem of emotion computing-emotion recognition. Results showed that it was able to facilitate short-term emotion recognition and its classification accuracy was above 78% in all cases [9]. Dong et al. designed a hierarchical attention network circuit for multimodal emotion computing, which used nanoscale memristors arranged in a cross-array configuration to construct a compact hierarchical attention network that can monitor the current user's mental health status. The research results indicated that this design helped achieve deep integration of neuromorphic electronics and mental health monitoring systems, which was conducive to designing a low-cost mental health monitoring system [10]. To solve the problems of high implementation cost and high power consumption in smart home monitoring systems, scholars such as Dong et al. proposed a multimode neural morphology sensation processing system based on memristor circuits. This system adopted a neural morphology multimode sensation processing module, which effectively collected and processed multiple sensory clues, accurately described the environment, and achieved instant response. This design not only promoted the progress of smart home applications, but also

reduced implementation costs, reduced power consumption, simplified device configuration, and was conducive to achieving carbon neutrality and carbon peak goals [11].

The aforementioned findings suggest that the identification of emotions is beneficial to enhance the analysis of learning behavior. The multidimensional feature fusion attention mechanism based on gaze detection and emotion recognition has good recognition and detection effects, which is conducive to detecting the attention situation of students in online teaching. To further the detection of students' attention in online teaching, this study will investigate face detection, human eye opening and closing detection, sight detection, and emotion recognition to enhance the effectiveness of attention detection.

3 Student attention detection in online teaching

To achieve student attention detection in online teaching, the study first uses the SDM algorithm to detect facial feature points and uses an improved least-squares ellipse-fitting algorithm to detect the open and closed state of the human eye. Then, it performs gaze detection and emotion recognition and constructs an attention detection system to achieve the detection of student attention in online teaching and learning. The study will introduce the Academic Emotion Recognition Index to improve attention detection methods, enabling them to detect students' academic emotions while their attention is focused, and then adjust their learning emotions in real time to judge their level of attention and fully leverage the advantages of online teaching.

3.1 Face feature points and human eye opening and closing detection in online teaching

The specificity of online teaching starts with the detection of whether the student is in front of the computer screen, i.e., the detection of whether the student has a human face or not. At the same time, there are currently more and more students wearing glasses, and in the detection, an algorithm that is not affected by glasses and is more robust and real time needs to be selected. In online teaching, it is necessary to check the attention level of students, so it is necessary to recognize, detect, and judge based on feature points. For example, facial contours such as eyes, mouth, and nose are a set of feature points that can depict the morphological structure of the face. Extracting feature points lays the foundation for subsequent research on gaze detection, emotion recognition, etc. [12]. Among them, the SDM is a representative regression analysis-based face feature point extraction method, which is a relatively simple calculation process, less sensitive to changes in light, has the characteristics of fast and stable, and is currently a popular face alignment method. To satisfy attention detection in online teaching, the research article adopts the SDM algorithm with many advantages to implement the detection of face feature points.

The eye is an important organ and contains a great deal of information that can be obtained from the movements and states of the eye as needed. One of the key factors affecting the extraction of visual information is the state of the human eye. In online teaching, human eye detection is an important tool to reflect students' attention status. The detection includes several indicators such as human eye opening and closing status, blink frequency, and gaze direction. In this study, the focus will be on detecting students' eye-opening and closing status. The study proposes an least square ellipse fitting (LSEF) to meet the requirements of human eye detection. After fitting the human eye contour accurately, if the ratio of the fitted ellipse area to the eye contour area is not within the threshold, the ellipse height can be corrected using the isometric method to obtain true eye width and height values. Then, the aspect ratio of the ellipse is used to determine whether the eye is in the open or closed state. Finally, the Canny operator in OpenCV is used to extract the contours of the eye region, and least squares are used to fit the ellipse so that it can quickly and accurately determine the relevant parameters of the human eye. When fitting the contours of the human eye region using the LSEF

method, failure to match eye contours is noted as a missing detection; even if a match is completed, the resulting contours may not be the true eye, or there may even be more than one. For this reason, it is necessary to improve LSEF to fit the eye contour more accurately, and to judge the eye conditions accordingly, so that it can obtain optimal results. The flow chart of the improved human eye opening and closing-state detection

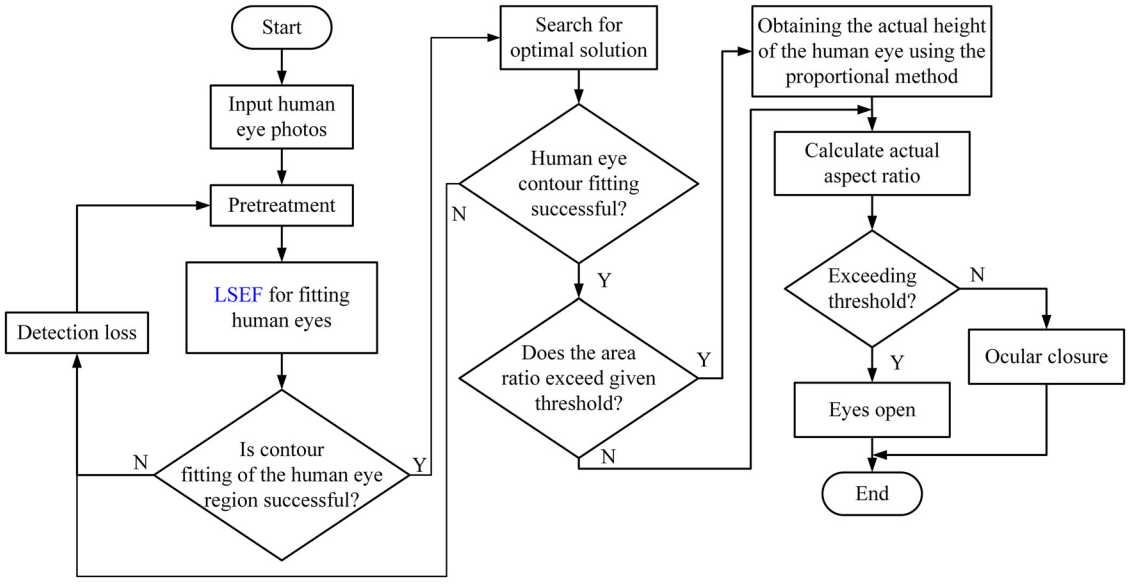

Figure 1: Flow chart of improved human eye opening and closing-state detection algorithm.

algorithm is shown in Figure 1.

In Figure 2, this algorithm consists of three parts: optimal solution search, threshold judgment, and actual eye aspect ratio. This method can effectively eliminate all kinds of interference, making it unique and able to fit eye contours better. Additionally, the improved algorithm can eliminate the influence of glasses on the target, so the detection of the target will be greatly improved even when glasses are worn [13,14]. If the ratio of the fitted ellipse area S to the eye contour area S_e is not in range, eye contour height should be calculated again and actual eye height can be found by the following equation:

$$\frac{S}{S_e} = \frac{a}{a_e}, \tag{1}$$

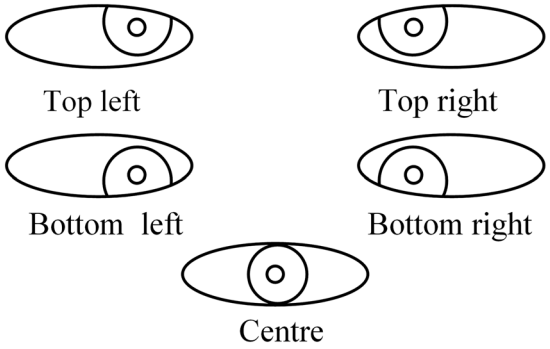

Figure 2: Sight landing situation.

where a and a_e represent the width of the human eye and eye contour, respectively. When the true contour height is obtained, the open and closed states can be determined directly from the width ratio p. In the case of $p > 0.24$, the eye is open, and in the opposite case, the eye is closed.

3.2 Sight detection methods in online teaching

The line of sight is an imaginary straight line between the eye and an object, which reflects information such as the relative position of the eye and the object. Line of sight detection and tracking technology is a very important human–computer interaction detection technology. Incorporating a line-of-sight detection module into online teaching can effectively improve students' emotional deficits and can facilitate the development of online teaching by adjusting the teaching mode according to the area of interest and length of stay of the learner's line of sight. The study will make use of an ordinary personal computers (PC's) low-resolution camera for detection and rational zoning of the sighting area. In the division of the line-of-sight landing area, it will be divided into five areas, namely, top left, bottom left, middle, top right, and bottom right. Based on the divided sight areas, the corresponding features are extracted and the students' sight drop on the screen is then analyzed [15]. By reducing the human eye contour to an ellipse and eliminating irrelevant information, it was possible to obtain six contour feature points, which were smoothly connected to obtain the human eye contour. The next step is to construct a coordinate system from the coordinates of the feature points and their position relations and $P_0(x_0, y_0)$ is the reference point. Combining the characteristics of online teaching and considering factors such as cost, the study selects an image gradient-based iris center localization algorithm for iris center extraction. It is assumed that in a circle, c' denotes the possible location of the pupil, c is the true circle center, g_i denotes the gradient vector of x_i, d_i is the displacement vector, and d_i and g_i are oriented in the same direction. The possible positions of the center of the circle are represented by x_i and $i \in \{1, \ldots, N\}$. When c is the true circle center, d_i and g_i have the same displacement vector; if not, there will be an angle between d_i and g_i, and then, the center of the circle is calculated as shown in the following equation:

$$c^* = \arg\max\left\{\frac{1}{N}\sum_{i=1}^{N}(d_i^T g_i)^2\right\}. \tag{2}$$

where c^* represents the center of the circle found. d_i and g_i are calculated as shown in the following equation:

$$\begin{cases} d_i = \dfrac{x_i - c}{\|x_i - c\|_2} \\ \forall i : \|g_i\|_2 = 1 \\ g_i = (\partial I(x_i, y_i)/\partial x_i, \partial I(x_i, y_i)/\partial y_i)^T. \end{cases} \tag{3}$$

In equation (3), $\partial I(x_i, y_i)$ represents the bias to x_i, y_i. However, when there is bright light or the influence of the eyelids, a bright pupil situation can occur, making the algorithm's localization biased. The study combines the luminance information by assigning a weight value w_c to c' by the pupil color being darker than the nearby area, and then, the final formula for the circle center is shown in the following equation:

$$c^* = \arg\max\left\{\frac{1}{N}\sum_{i=1}^{N} w_c (d_i^T g_i)^2\right\}. \tag{4}$$

The algorithm detects the center of the iris with good results when the test subject has a small head shift and wears glasses. However, because the algorithm is also influenced by eyebrows, for example, the iris center can be positioned outside the eye contour, and the large area of the detection area makes the calculation process more complex. Therefore, the study uses the human eye contour extracted using the SDM algorithm as the detection center region of the iris and uses $P_r(x_r, y_r)$ to represent the iris center point. In turn, it is possible to obtain the learner's visual dropout through the visual region division and $P_r(x_r, y_r)$, $P_0(x_0, y_0)$, as shown in Figure 2.

Figure 2 shows the five sightline situations, from left to right and from top to bottom: top left, top right, bottom left, bottom right, and center. Considering that human eye movements have the same or opposite pattern, this study will take the right eye as an example to investigate the sighting situation, and the plane distance between $P_r(x_r, y_r)$ and $P_0(x_0, y_0)d$ can be found by the following equation:

$$d = \sqrt{(x_0 - x_r)^2 + (y_0 - y_r)^2}. \tag{5}$$

However, in practice, the detection accuracy decreases significantly when the detector's line of sight looks downward, and the image gradient-based iris center detection algorithm also suffers from reduced iris centroid extraction accuracy when the line of sight is directed downward, so there is a need to find an effective method that can reduce the influence of the eyelid on the detection results. To address these issues, an improved line-of-sight detection method is proposed, which introduces an eye aspect ratio parameter based on the original line-of-sight detection method p. The process of the improved line-of-sight detection method is shown in Figure 3.

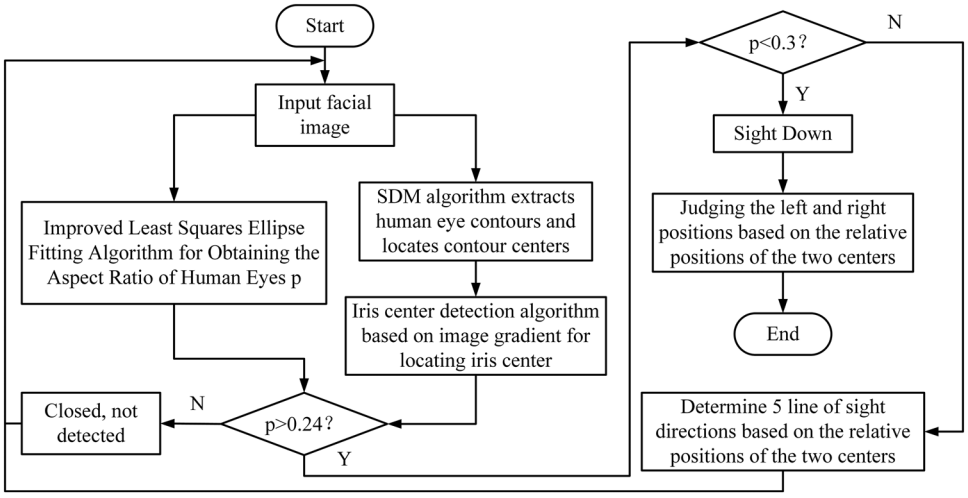

Figure 3: Improved line of sight detection method.

In Figure 3, it can be seen that p is mainly used to determine whether the human eye is closed; when $0.24 < p < 0.3$, this implies that when the detector gazes downward, the algorithm can assess the left and right directions, distinguishing between left-down and right-down gaze orientations. Through this enhanced gaze detection method, it is possible to effectively and accurately detect the specific landing position of the gaze.

3.2.1 Fuzzy reasoning-based emotion recognition and attention detection system

The attention detection system, including indicators such as face presence detection, human eye opening and closing detection, and sight fall detection, significantly improves the human–computer interaction experience in online teaching environments. It also enables the real-time detection of learners' attention status, making a valuable contribution to the development of online education [16,17]. However, analysis of students' attentional status from these perspectives alone does not provide a true understanding of students' attentional status. Based on this, the study added emotion recognition to the system, hoping to improve the attention detection system and make the most of the advantages of the online teaching model. However, emotion recognition has subjectivity, and fuzziness, making traditional emotion recognition methods difficult to effectively handle. Fuzzy theory can simulate the human mind and quantify the subjective concepts of human beings [18]. Fuzzy theory converts the problem of subjectivity and uncertainty into a problem of affiliation for

quantitative study. Based on this principle, a new fuzzy reasoning method is proposed. The fuzzy reasoning method is applied to analyze students' emotions qualitatively and quantitatively. The process of emotion identification based on fuzzy reasoning is shown in Figure 4.

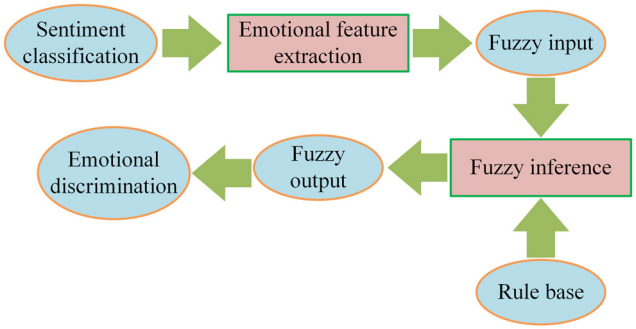

Figure 4: Emotion recognition process based on fuzzy reasoning.

In Figure 4, emotions are first classified, then emotional features are selected, fed into fuzzy theory, fuzzy reasoning is performed, and the results of fuzzy reasoning are output in combination with a rule base, which, in turn, enables objective identification of emotional states. Considering the characteristics of online teaching, this study focuses on the learners' facial emotions and classifies learning emotions into three types: pleasant, bland, and distressed. The external representation of facial emotion change is the change in features of the parts of the face associated with it, which are extracted and analyzed to identify the emotional state of the learner while learning [19]. The mouth and eye regions are the most obvious areas of emotional change, so the parameters of top and lower mouth curvature, eyebrow length, and forehead texture features will be extracted and analyzed. In the process of mouth angle curvature extraction, it is assumed that A and B represent the left and right mouth angle points, respectively, while M_1 and M_2 are the center points of the top and lower lips, respectively. The SDM algorithm is used to extract these four feature points, and it is assumed that $A(x_a, y_a)$, $B(x_b, y_b)$, $M_1(x_{M_1}, y_{M_1})$, and $M_2(x_{M_2}, y_{M_2})$ represent the coordinates of A, B, M_1, and M_2, respectively, and S_1 and S_2 represent the tangent values of the top and lower curvature of the mouth angle, respectively, in the following equation:

$$\begin{cases} S_1 = \left| \dfrac{y_{M_1} - y_a}{x_{M_1} - x_a} \right| + \left| \dfrac{y_{M_1} - y_b}{x_{M_1} - x_b} \right| \\ S_2 = \left| \dfrac{y_{M_2} - y_a}{x_{M_2} - x_a} \right| + \left| \dfrac{y_{M_2} - y_b}{x_{M_2} - x_b} \right| \end{cases}. \tag{6}$$

After calculating S_1 and S_2 by equation (6), it can facilitate the recognition of emotions. The SDM algorithm can extract three learners' eyebrow features in different emotional states d_1, d_2, and d_3, which, in turn, allows the calculation of the overall length of the eyebrows d_{mm}, as shown in the following equation:

$$d_{mm} = d_1 + d_2 + d_3. \tag{7}$$

There is a connection between the forehead region and the eyebrows, and the skin folds shown in the forehead region can vary in different emotional states. Therefore, the texture of the forehead can be detected to reflect the changes in the forehead skin folds and thus identify the emotional state [20]. The texture information of the forehead region is detected using the Canny operator that comes with OpenCV. The fuzzy sets extend the affiliation from 0 and 1 to [0, 1], thus representing the attribution level of each element in the set that is consistent with certain requirements. Assuming that U represents a set and $\{U\}$ is an element in U, the fuzzy set can be represented using the affiliation function. Assuming that A^* is a fuzzy subset of U and u_A is a representation of the affiliation function, the mapping relationship can be derived from the following equation:

$$u_{A^*} : U \to [0, 1]. \tag{8}$$

The subordination of U to A^* is $u_{A^*}(u)$, which describes the degree to which the elements in U belong to A^* and takes the value range [0, 1]. When the value of $u_{A^*}(u)$ is closer to 1, the element belongs to set uA^* to a greater extent; conversely, when the value of $u_{A^*}(u)$ is closer to 0, it means that element u belongs to the set A^* to a lesser extent. After each module of visual attention detection, it can be fused to realize the multidimensional feature fusion attention detection based on face detection, human eye opening and closing detection, visual detection, and emotion recognition in online teaching, and the flow chart is shown in Figure 5.

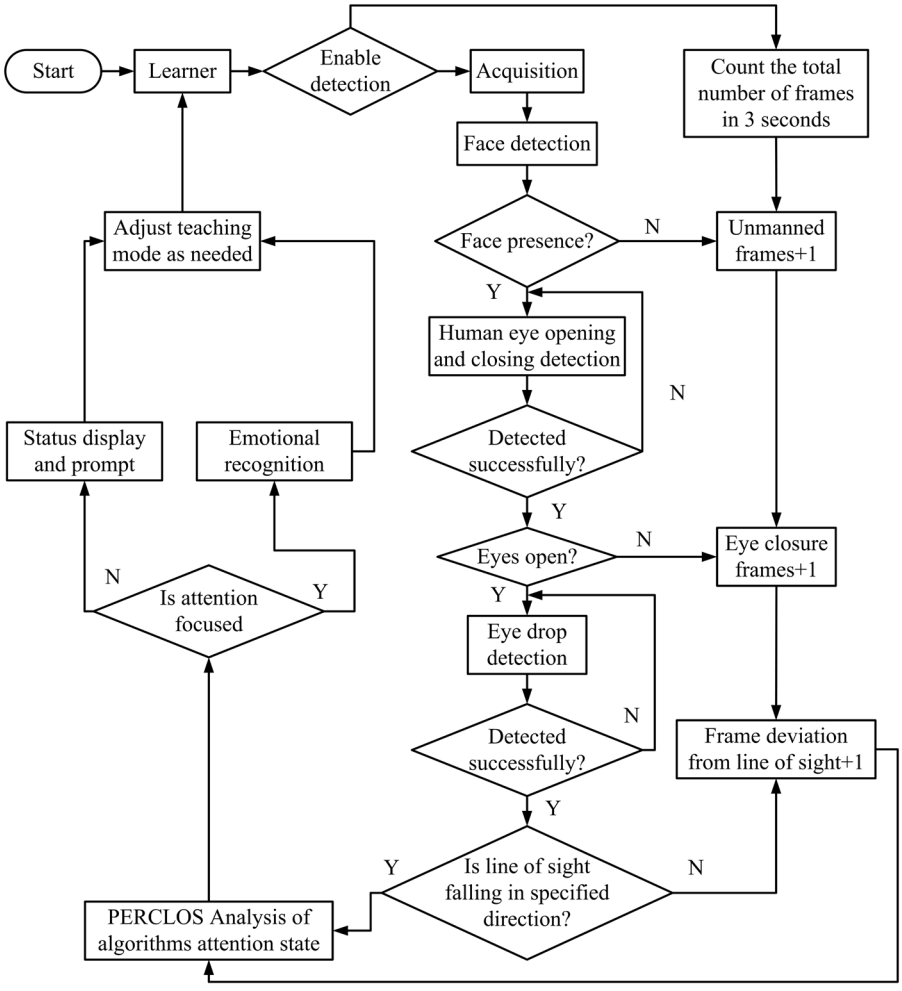

Figure 5: Flow chart of attention detection method.

In the attention detection in Figure 5, the discrimination of learners' attention status is achieved by fusing multiple face features, thus solving the problem that cannot be detected by a single feature. Specifically, the study will introduce an academic emotion recognition index to improve the attention detection method so that it can detect students' academic emotions while they are paying attention, and thus make real-time adjustments to their learning emotions, to make judgments about their attention levels to fully exploit online teaching.

4 Analysis of multidimensional feature fusion attention mechanisms in online teaching

A comprehensive investigation was conducted on the multidimensional feature fusion attention mechanism in online teaching. This analysis involved the examination of multiple factors, including face feature point detection results, line of sight detection results, emotion recognition results, and attention detection results. By integrating these facets into the analysis, the research aimed to validate the accuracy of the proposed method for detecting attention in online teaching contexts.

4.1 Analysis of face feature points and line of sight detection results

The study focused on capturing real-time facial images of students using a typical PC equipped with a low-resolution camera. Attention detection methods were implemented using software tools, such as Visual Studio 2013. The Intel Core i7-6700HQ had a 4-core CPU, 8GB of memory, and a camera pixel size of around 300,000. The SDM algorithm set thresholds and scales for facial detection to effectively detect facial feature points in low-resolution cameras, optimizing the detection speed and accuracy of feature points to adapt to real-time video stream processing. The least-squares ellipse-fitting algorithm-adjusted algorithm parameters to adapt to accurate detection of human eye opening and closing status in low-resolution images, including parameters such as eye size and position and optimized the algorithm to handle eye state detection under different lighting and angle conditions. Eye detection and emotion recognition required determining key features such as eye direction and pupil size for eye detection. The study involved selecting and optimizing emotion recognition algorithms, including facial expression-based emotion recognition, along with related model parameters and training datasets. The research identified key characteristics of academic emotions and designed corresponding emotion recognition indicators. Emotion recognition algorithms were carefully selected and optimized to ensure accurate recognition of academic emotions. Attention detection methods were combined with academic emotion recognition for model improvement by adjusting model weights and feature selection, enhancing the ability to judge student attention and emotions. Adjustment strategies and mechanisms were designed to achieve real-time adjustment of student learning emotions, considering the potential adjustment amplitude and methods under different emotional states.

The faces were detected and 70 key feature points were obtained using the SDM algorithm for face feature point extraction. To analyze its performance, 20 online students were selected and asked to sit in front of a computer screen and change and adjust the direction of their eyes during the experiment in the order of top left, bottom left, middle, top right, and bottom right. During this process, the experimental students needed to maintain their heads stationary or slightly rotate, and recorded a 1-minute video for each student, each recording 10 videos, including 5 groups each with and without glasses. The videos of the 20 students were examined and analyzed, and the results of the tests of three randomly selected students are shown in Figure 6.

In Figure 6, the accuracy of detecting eye state for students with and without glasses was compared using a modified least-squares Nan's circle-fitting algorithm. The detection accuracies for Student 1 with and without glasses were 82.8 and 91.7%, respectively; for Student 2, 86.6 and 93.1%; and for Student 3, 84.1 and 91.2%. That is, the average accuracy rates with and without glasses were 84.5 and 92.0%, respectively. This indicates that under the same conditions, the measurement accuracy without glasses was higher due to lens reflection. To further analyze the effectiveness of the eye open/closed-state detection algorithm, the study selected two common human eye state detection methods for comparison and analyzed the detection accuracy and the average detection time taken by the three algorithms with and without glasses, and the results are shown in Table 1.

In Table 1, the detection accuracies of the three algorithms are 74.8, 82.3, and 84.5% with glasses on and 88.7, 90.7, and 92.0% without glasses on, respectively; the average time consumption cases of the three algorithms are 13, 103, and 17 ms with glasses on and 12, 100 and 15 ms without glasses on, respectively. This indicates that the study method has high detection accuracy and fast real-time performance. Afterward,

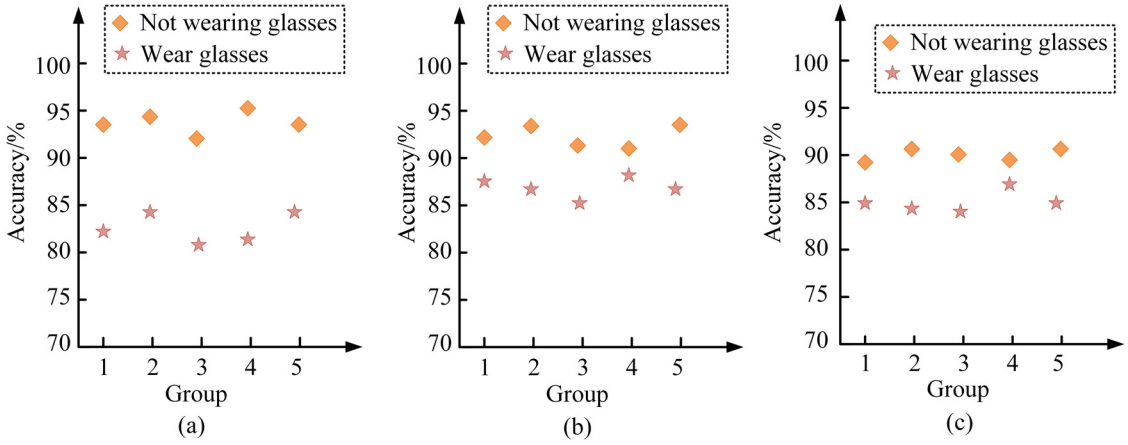

Figure 6: Eye opening and closing status detection results of three students: (a) Volunteer 1, (b) Volunteer 2, and (c) Volunteer 3.

Table 1: Detection accuracy and average detection time of three algorithms with and without glasses

Project	Whether to wear glasses	Comparison method 1	Comparison method 2	Research method
Accuracy	Wear glasses	74.8%	82.3%	84.5%
	Without glasses	88.7%	90.7%	92.0%
Average time consumption	Wear glasses	13 ms	103 ms	17 ms
	Without glasses	15 ms	100 ms	15 ms

five students were selected from a video of 20 students, and their visual directions were analyzed and compared with the improved visual detection method, and the results are shown in Figure 7.

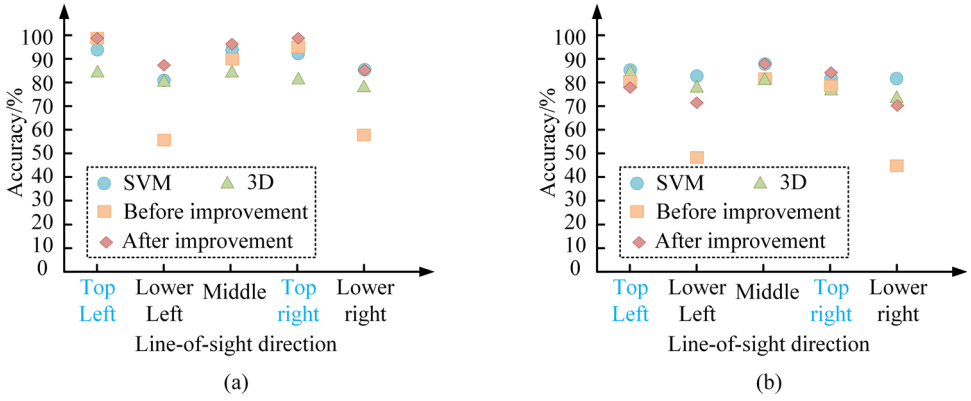

Figure 7: Accuracy of each algorithm in different line of sight directions: (a) accuracy of various algorithms in different line of sight directions without wearing glasses and (b) accuracy of various algorithms in different line of sight directions when wearing glasses.

In Figure 7, the accuracy of learners in both the top-left and top-right views was good, with the highest accuracy of 99.0% in both views and 97.3% in the middle view. The accuracy rates for the bottom-left and bottom-right detections were 58.3 and 60.9%, respectively, far short of the recognition requirements. In the three visual directions of top left, top right, and middle, the accuracy of recognition using the improved method was the same as before the improvement. While in the lower-left and lower-right directions, the accuracy was up to 90% in the lower-left direction and 86% in the lower-right direction, both of which were

greatly improved and fully met the test requirements. When they put on the glasses, the prediction accuracy also improved significantly, with the highest accuracy of 73.9 and 72.3% in the lower-left and lower-right directions. Results showed that the improved method approach had high recognition accuracy and was more accurate compared to the existing methods.

4.2 Analysis of emotion feature extraction results and attention detection results

There is a certain connection between the forehead region and the eyebrows, and the skin folds shown in the forehead region can vary in different emotional states. Therefore, the detection of the forehead texture can reflect the changes in the forehead skin folds, and thus, the emotional state can be identified. The forehead texture was first detected using different Canny operators, and the results are shown in Figure 8.

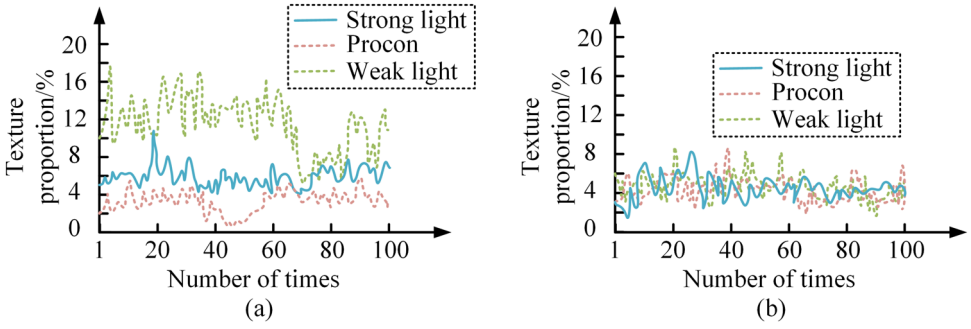

Figure 8: Detection results of forehead texture using different Canny operators: (a) traditional canny operator and (b) adaptive canny operator.

Figure 8 shows that in the traditional Canny operator, the percentage of forehead texture for the same emotion under different lighting conditions of strong light, normal light, and low light had a large difference; while in the adaptive Canny operator, the percentage of forehead texture for the same emotion under different lighting was relatively close. Since the forehead texture ratio also changed when the distance between the learner and the screen was changed, it is assumed that the lighting conditions were normal lighting, the face area was judged, and a threshold was set to obtain the forehead texture ratio detection results under different distances, as shown in Figure 9.

In Figure 9, the normal distance is between 50 and 70 cm from the computer screen, and the far distance is between 85 and 105 cm from the computer screen. In Figure 10, after setting the lighting conditions and

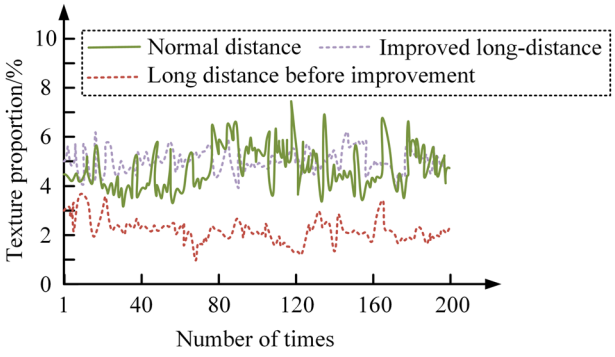

Figure 9: Frontal texture proportion detection results at different distances.

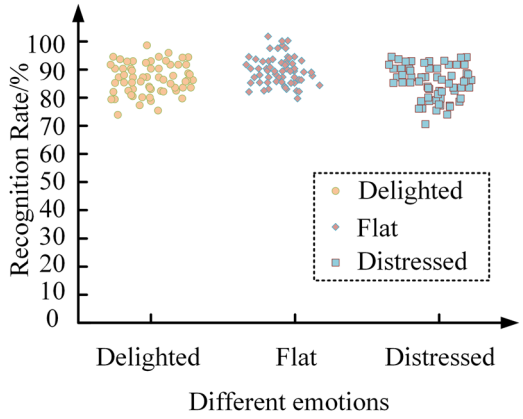

Figure 10: Results and recognition rate of emotion recognition.

thresholds, it was able to provide a better improvement of the forehead texture changes caused by the distance variation. The emotional changes of the learners were then identified and analyzed using an emotional recognition method based on fuzzy inference, and the learning emotions were classified into three types: pleasant, bland, and distressed, and information such as the percentage of relevant textures was input into the fuzzy inference. The relevant data of five learners, each involving 100 data, were taken, and the abnormal data were eliminated to obtain the results and recognition rate of emotion recognition, and the results are shown in Figure 10.

In Figure 10, the results show that the research method achieved an average recognition rate of 90.7% for the three different types of learning emotions, which fully meets the requirements for learning emotion recognition in online teaching. Relatively speaking, the accuracy rate was somewhat higher under the bland emotion, at 92%. Under pleasant emotions, the accuracy rate reached 91%, and for distressed emotions, there was an 89% accuracy rate. When the forehead texture percentage reached 8% or more, its detection rate reached 100%. Afterward, the facial images of the learners were captured in real time using a normal PC with its camera, and the learners were kept at a normal distance from the computer screen to capture learning videos of 15 learners, each video being 10 minutes long and detected 200 times for each segment. The attentional state during the learning process was detected and analyzed in Figure 11.

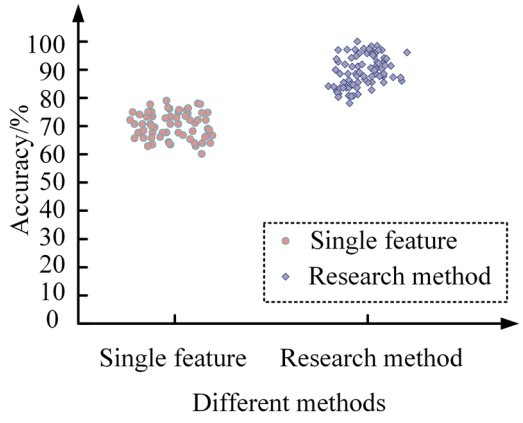

Figure 11: Comparison of detection results of attention states using different algorithms.

In Figure 11, the use of a single feature to detect a student's attentional state was not ideal. When the eyes were open and the line of sight was away from the specified direction, the single human eye opening and closing detection did not provide an accurate judgment of the results, and a missed and false detection

occurred. The use of the research algorithm, which performed attention detection on the learner, incorporated both human eye open/closed detection and line of sight deviation detection, allowing for effective improvement on these issues and a significant increase in accuracy. The detection accuracies for the single feature and the research method were 74.1 and 91.9%, respectively.

5 Conclusion

As computer technology develops and in the context of the post-epidemic era, online teaching has become a mainstream teaching method, but due to the lack of face-to-face communication, teachers are unable to make judgments on students' learning emotions and attention situations, as well as to make real-time adjustments to teaching strategies. To this end, the study proposes a multidimensional feature fusion attention machine that incorporates face feature point detection, human eye open/close-state detection, line of sight detection, and emotion recognition, and improves the detection method. Results showed that the average accuracy of detection using improved least-squares nanocircle-fitting algorithm was 84.5 and 92.0% for students with and without glasses, respectively, and under the same conditions, the measurement accuracy without glasses was higher. The highest accuracy was 90% in the lower-left direction, 86% in the lower-right direction, and 73.9 and 72.3% in the lower-left and lower-right directions, respectively. With set lighting conditions and thresholds, it was able to provide a good improvement in forehead texture changes caused by changes in distance and proximity. In a flat mood, the accuracy was a little higher at 92%. When forehead texture accounted for 8% or more, its detection rate was up to 100%. The detection accuracies for the single feature and the study method were 74.1 and 91.9%, respectively. This indicates that the research method has high detection accuracy and fast real-time performance with high recognition accuracy. This study provides an effective tool for monitoring student attention and emotions in online teaching environments, which helps teachers to understand students' learning status in real time and improve teaching effectiveness. Through accurate attention and emotion detection, teachers can adjust teaching methods and content based on students' real-time reactions, achieving more personalized and efficient teaching. In the post-pandemic era, with the popularization of online teaching, this technology provides an innovative way to solve the problems of student participation and lack of teacher feedback, enhancing the interactivity and effectiveness of online teaching. There are still some shortcomings in this study, such as only dividing the gaze direction into five types, but cannot fully summarize the student's observation direction. Therefore, future research will improve the division of gaze areas to better detect student attention.

Funding information: The authors state no funding involved.

Author contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Yurong Huang and Guang Yang. The first draft of the manuscript was written by Yurong Huang and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflict of interest: The authors report there are no competing interests to declare.

Data availability statement: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

References

[1] Cui Y, Han G, Zhu H. A novel online teaching effect evaluation model based on visual question answering. J Internet Technol. 2022;23(1):91–8.

[2] Dutta GP. Identifying challenges and opportunities in teaching chemistry online in India amid COVID-19. J Chem Educ. 2021;98(2):694–9. J Chem Educ. 2021;98(2):694–9.
[3] Liu S, Zhang M, Fang M, Zhao J, Hou K, Huang CC. Speech emotion recognition based on transfer learning from the FaceNet framework(a)). J Acoust Soc Am. 2021;149(2):1338–45.
[4] Cui Y, Wang S, Zhao R. Machine learning-based student emotion recognition for business English class. Int J Emerg Technol Learn (iJET). 2021;16(12):94–107.
[5] Huang D, Zhang WX. Research on learning state based on students' attitude and emotion in class learning. Sci Program. 2021;2021(Pt. 12):99441761-9944176.11.
[6] Wang S. Online learning behavior analysis based on image emotion recognition. Traitement du Signal. 2021;38(3):865–73.
[7] Wu J, Chen B. English vocabulary online teaching based on machine learning recognition and target visual detection. J Intell Fuzzy Syst. 2020;39(2):1745–56.
[8] Xuan Z, Liu C, Zhang Y, Li Y, Kang Y. A brain-inspired ADC-free SRAM-based in-memory computing macro with high-precision MAC for AI application. IEEE Trans Circuits Syst II: Express Briefs. 2022;70(4):1276–80.
[9] Tan C, Arlija M, Kasabov N. NeuroSense: Short-term emotion recognition and understanding based on spiking neural network modelling of spatio- temporal EEG patterns. Neurocomputing. 2021;434(Apr. 28):137–48.
[10] Dong Z, Ji X, Lai CS, Qi DL, Zhou GD, Lai LL. Memristor-based hierarchical attention network for multimodal affective computing in mental health monitoring. IEEE Consum Electron Mag. 2023;12(4):94–106.
[11] Dong Z, Ji X, Zhou G, Gao M, Qi D. Multimodal neuromorphic sensory-processing system with memristor circuits for smart home applications. IEEE Trans Ind Appl. 2022;59(1):47–58.
[12] Jayasimha Y, Reddy RVS. A robust face emotion recognition approach through optimized SIFT features and adaptive deep belief neural network. Intell Decis Technol. 2019;13(3):379–90.
[13] Gupta D, Bansal P, Dr. kavita. Emotion recognition: Differences between spontaneous dialogue and active dialogue. J Shanghai Jiaotong Univ (Sci). 2021;16(9):633–44.
[14] Bota PJ, Wang C, Fred ALN, Silva HP. A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. IEEE Access. 2019;7(99):140990–1020.
[15] Ordonez CE, Blotta E, Pastore JI. Isophote-based low-computing-power eye-detection embedded-system. IEEE Lat Am Trans. 2020;18(2):336–43.
[16] Ntalampiras S. Speech emotion recognition via learning analogies. Pattern Recognit Lett. 2021;144(Apr):21–6.
[17] Ocquaye ENN, Mao Q, Xue Y, Song H. Cross lingual speech emotion recognition via triple attentive asymmetric convolutional neural network. Int J Intell Syst. 2021;36(1):53–71.
[18] Xiao C, Cai H, Su Y, Shen L. Online teaching practices and strategies for inorganic chemistry using a combined platform based on dingtalk, learning@ ZJU, and wechat. J Chem Educ. 2020;97(9):2940–4.
[19] Alsemawi MRM, Mutar MH, Ahmed EH, Hanoosh HO, AI- Dhief FT. Emotions recognition from human facial images using machine learning technique. Solid State Technol. 2020;63(5):8749–61.
[20] Guo Y, Mustafaoglu Z, Koundal D. Spam detection using bidirectional transformers and machine learning classifier algorithms. J Comput Cognit Eng. 2022;2(1):5–9.