Fuzzy-Based Head Attitude Estimation for Improved Students' Concentration Evaluation

Zheng Zhou[1[0009−0001−3981−6328]], Kaoru Hirota[1[0000−0001−5347−6182]], Yaping Dai[1[0000−0001−8795−5333]], MD SHARIFUL ISLAM[1[0009−0009−1120−6278]], Bemnet Wondimagegnehu Mersha[1[0000−0002−1000−8984]], Wei Dai[2[0009−0002−1397−451X]], and Yumin Lin[2[0009−0004−6617−8378]]

[1] Beijing Institute of Technology, Beijing 100081, China
[2] River Security Technology Co., LTD, Shanghai 200336, China
3220220886@bit.edu.cn

Abstract. In order to evaluate students' concentration in offline education, an algorithm based on fuzzy comprehensive evaluation is proposed. The algorithm evaluates students' concentration by measuring their head attitude angle, which consists of three modules: face key points detection, head attitude angle measurement and concentration decision, and outputs the curve of students' overall concentration score over time. Compared with other concentration evaluation methods, the proposed algorithm achieves the evaluation of overall students' concentration under low pixel video and is suitable for most offline classrooms with monitoring devices. The overall functional effectiveness of the algorithm was tested with a classroom video dataset of 35 students. The algorithm outputs students' concentration scores at 30 FPS, meeting the requirement of a real-time classroom. The algorithm's scores were compared to the artificial scores of 15 experts, resulting in an average accuracy of 88.3% and a Pearson's correlation coefficient of 0.936 between the two, thus validating the effectiveness of the algorithm. The proposed algorithm can provide educators with a reference for educational effectiveness and help to realize the automatic assessment of educational quality in the future.

Keywords: Offline education · Concentration evaluation · Head attitude estimation· Fuzzy comprehensive evaluation.

1 Introduction

For students in offline education, concentration reflects mastery of teaching content. Some studies have analyzed videos or images through computer vision methods to detect concentration. Among them, Savchenko utilizes a multi-task convolutional neural network (MTCNN) for face detection and employs a trained lightweight convolutional neural network (CNN) to classify images and obtain facial sentiment features. This process outputs predicted levels of concentration and individual moods [1]. Pabba uses a methodological framework for training facial expression recognition models using CNNs to recognize students' facial expressions from videos and classify students' concentration into low, medium, and high levels [2]. Thomas uses computer vision technology to analyze the students' concentration levels from their facial expressions, head postures, and eye gaze and uses machine learning algorithms to make decisions that categorize the level of concentration as attentive or distracted [3]. Zhong scores head attitude based on the left and right rotation angle of the head and the head-up and head-down angle. Fatigue is scored through the results of eye and mouth closure detection, and emotion is scored in combination with the results of facial expression detection. The scores from these three aspects are combined to quantify the degree of concentration [4]. Most methods evaluate concentration status by identifying students' facial expressions or the degree of eye and mouth closure. However, there are some shortcomings in current students' concentration research.

- Most of the studies are based on single-person concentration concentration, i.e., only one student is present in a single image frame. However, there are a large number of students in real classroom scenarios and fewer studies exploring the overall concentration of multiple students in a single image frame.
- The low resolution of video provided by the surveillance equipment in some classrooms makes it difficult to analyze eyes and mouths. Additionally, the changes in students' expressions in the classroom are not obvious, rendering the study of facial expressions less meaningful.

Therefore, in responding to these issues, a concentration evaluation algorithm based on fuzzy comprehensive evaluation is proposed to evaluate students' classroom concentration by measuring their head attitude angle. By analyzing the classroom surveillance video, we can measure the head attitude angle. The algorithm can be applied to most classrooms with cameras. The overall concentration in the classroom better reflects the effectiveness of the teacher's teaching, facilitating the teacher's scheduling of the lesson. However, in classrooms with many students, it becomes challenging for the teacher to detect the overall concentration state of the students. Additionally, students' concentration in the classroom is a vague concept without a clear definition. To address these issues, the proposed algorithm detects students' head attitude angle. Using the detected head attitude angle as input, the student's concentration score output is obtained through a comprehensive fuzzy evaluation. The student's concentration scores in the classroom from the algorithm provide a reference for teachers' teaching effectiveness.

Section 2 describes the general framework of the concentration evaluation algorithm. Section 3 describes experimental validation and analysis of results.

2 General Framework of Concentration Evaluation Algorithm

In this section, the general structure of the algorithm is presented and the theoretical approach of the three modules is described.

2.1 Framework of Algorithm

The experimental dataset consists of two parts: a single-person video dataset captured by the computer's front-facing camera and a classroom video dataset with 35 students. The overall flow of the algorithm is shown in Fig. 1. Taking the classroom video as input, the proposed concentration evaluation algorithm consists of three modules: face key points detection, head attitude angle measurement and concentration decision. In the face key points detection module, the Retinaface algorithm is used to extract students' face images from the classroom surveillance video image and detect the 2D key point coordinates of each face image. In the head attitude angle measurement module, the 2D key point coordinates of the face are matched with the 3D face model to derive the rotation matrix. The head attitude Euler angle is solved by the rotation matrix. In the concentration decision module, the head attitude angle of each person is evaluated as the corresponding level through a fuzzy comprehensive evaluation. The concentration levels of all detected students are fused to obtain the students' concentration scores in the classroom. The curve of the students' concentration scores over time is the output.

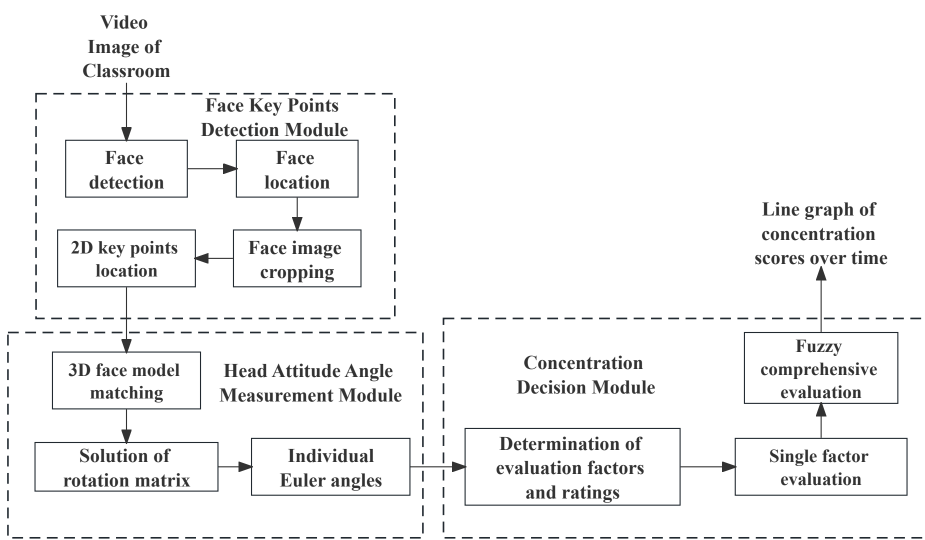

Fig. 1. The overall flow of the algorithm

2.2 Face Key points Detection Module

The algorithm uses the Retinaface model for face key points detection. Use the center of left eye, the center of right eye, the tip of the nose, the left corner of the mouth, and the right corner of the mouth as the five key points of the face to observe. This module detects the face in the video image and locates the five key points.

Retinaface is a face detection model proposed by the Insight Face team in 2019 [5]. The network structure consists of four modules in total: the backbone network, feature pyramid, context module, and multitask loss function.

The method uses the Residual Network ResNet50 as the backbone network with feature pyramid layers from P2 to P6 [6]. Here, P2 to P5 computes the outputs of the corresponding ResNet residual phases (layers C2 to C5) by utilizing top-down and cross-connections. Separate context modules are also applied to these five feature pyramid layers to improve the sensory field and increase the capability of contextual modeling. Retinaface's multitasking loss function corresponds to four parallel branches, including the face classification loss function, the detection of face frame position regression function, the key point positions regression function, and the face pixel vs. 3D position and correspondence. Additionally, in the Loss Supervised Signal, RetinaFace focuses on increasing the importance of bounding box and key points location. The module output consists of three parts: the probability of the face image, the coordinates of the face frame, and the coordinates of the five face key points.

2.3 Head Attitude Angle Measurement Module

The video image represents the 2D planar world and the head attitude angle is measured for the 3D world. The 3D head attitude angle of the student needs to be restored from the 2D key point information obtained in Section 2.2. The 3D stereo image can be projected to obtain a 2D planar view, and the rotation angle of the transformation can be obtained by solving the rotation matrix. The problem of detecting the 3D head attitude angle of a student from a video frame is transformed into the mapping and calibration between four coordinate systems: the 2D-pixel coordinate system, the 2D image coordinate system, the 3D camera coordinate system, and the 3D world coordinate system [6]. The transformation relationship of the four coordinate systems [7] is shown in

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \tag{1}$$

where $(u,\ v)$ are the 2D coordinates of the key points of the face in the pixel coordinate system detected in Section 2.2, and $(X_w,\ Y_w,\ Z_w)$ are the 3D standard coordinates in the world coordinate system. $(f_x,\ f_y)$ are the components of the focal length in the horizontal and vertical axes, and $(u_0,\ v_0)$ are the optical center of the camera. The rotation translation matrix $(R,\ T)$ is the target object attitude matrix, where R denotes the rotation matrix, and T denotes the translation matrix.

Consider that the focal length is approximated as the width of the image, the center of the image is close to the optical center, and there is no radial distortion of the lens. Find the desired rotation matrix R.

The pixel coordinates of the five key points were obtained using the Dlib module and referenced to the standard frontal facial features to obtain the 3D standard reference coordinates of the key points, as shown in Table 2. The algorithm uses three Euler angles, Pitch, Yaw and Roll, to describe the head attitude, representing the student's head down and up, head turn left and right, and head tilt left and right respectively [8].

The rotation matrix R is shown in

$$\begin{aligned}\mathbf{R} &= \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \\ &= \begin{bmatrix} \cos\alpha\cos\beta & \cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma \\ -\sin\beta & \cos\beta\sin\gamma & \cos\beta\cos\gamma \end{bmatrix},\end{aligned} \tag{2}$$

it is converted to Euler angles using

$$\begin{cases} \gamma = \arctan\frac{r_4}{r_1} \\ \alpha = \arctan\frac{-r_7}{r_1\times\cos\gamma + r_4\times\sin\gamma} \\ \beta = \arctan\frac{r_3\times\sin\gamma + r_6\times\cos\gamma}{r_5\times\cos\gamma - r_2\times\sin\gamma} \end{cases}, \tag{3}$$

with γ, α, and β representing Roll eagle, Yaw eagle, and Pitch eagle, respectively [9].

The Euler angle is the angle between the transformed straight line and the 2D image straight line, which may differ from the real head attitude Euler angle by 180° (e.g., 2° is on the same straight line as -178°). Since there is a limit to the human head rotation, when the head rotation is too large, Retinaface cannot detect it in the face image. We limit angle range to [-90°, 90°] using

$$\theta = arcsin(sin\theta), \tag{4}$$

where θ is the Euler angle.

The camera is located high in the front of the classroom. When students are sitting upright facing the camera, concentration is in a full mark state. The Euler Angle in this state is defined as a forward gazing value. Combining the principle of the head attitude measurement module with the experimental verification of the surveillance video, the angular range and forward gazing values of the Euler angle are detected and obtained, as shown in Table 1.

Table 1. Angular range and standard values of Euler angles

Euler angle	Detection range	Forward gazing value
Pitch	[-30,30]	-10
Yaw	[-60,60]	0
Roll	[-48,48]	0

Table 2. 3D standard reference coordinates for face key points

Key point	Reference coordinates
Center of the left eye	-225,170,-135
Center of the right eye	225,170,-135
Tip of the nose	0,0,0
Corner of the left mouth	-150,-150,-125
Corner of the right mouth	150,-150,-125

2.4 Concentration Decision Module

After obtaining the Euler angle of each student, data fusion through fuzzy comprehensive evaluation is performed to obtain the students' concentration. Data fusion methods include Bayesian theory [10], Kalman filter [11], etc. The reasons for choosing fuzzy comprehensive evaluation are as follows:

- Concentration is an indicator that is not easy to judge and is a fuzzy objective. The concentration status can be quantified into a specific score by using fuzzy methods.
- There is some error in head attitude measurement. The use of fuzzy methods can reduce the effect of error on the evaluation of concentration.
- A single-factor ranking using Euler angles as an evaluation factor is consistent with the fuzzy approach.

Therefore, we choose fuzzy comprehensive evaluation to fuse students' concentration compared with other data fusion methods [12]. Firstly, the evaluation factors and evaluation sets are determined. Three Euler angles (Pitch, Yaw and Roll) are selected as the first evaluation factors, and different ranges of angle values are chosen as the second evaluation factors. The evaluation factors are delineated in Table 3. Based on the student's concentration in the classroom, the concentration evaluation set is: V={ v_1(very focused),v_2(focused),v_3(unfocused),v_4 (very unfocused)} . The evaluation sets are outlined in Table 4. For each student's set of Euler angles ($pitch$, yaw, $roll$), the first-level factors are determined by Table 3. The corresponding level set of Euler angle ($level_p$, $level_y$, $level_r$) is obtained. Then, the evaluation rules in Table 4 determine the evaluation set to which the Euler angles set belongs. If $max(level_p,level_y,level_r)= i$, the evaluation level is v_i.

Then, the weight values of the three Euler angles are determined. The hierarchical analysis method proposed by Thomas L. Saaty, a member of the American Academy of Engineering in 1971, is mainly applied to decision-making problems with multiple uncertainties [13]. Hierarchical analysis is used to determine the weights of the second level of factors (Euler angles). Pitch reflects the students' head-down and head-up states, yaw reflects the head-turning variations, and roll reflects the crooked head variations. In the actual classroom, students' head attitude changes are more often shown as pitch and yaw changes. Based on the

Table 3. Division of evaluation factors

First factors	Second factors	
Name of factor	Name of factor	Level
Pitch	[-15,-5]	1
	[-20,-16]∪[-4,0]	2
	[-25,-21]∪[1,5]	3
	[-30,-26]∪[6,30]	4
Yaw	[-15,15]	1
	[-30,-16]∪[16,30]	2
	[-45,-31[∪[31,45]	3
	[-60,-46]∪[46,60]	4
Roll	[-12,12]	1
	[-24,13]∪[13,24]	2
	[-36,-25]∪[25,36]	3
	[-48,-37]∪[37,48]	4

Table 4. Division of evaluation sets

Evaluation sets	Level	Score	Evaluation rules
Very focused	v_1	S_1	$max(level_p, level_y, level_r) = 1$
Focused	v_2	S_2	$max(level_p, level_y, level_r) = 2$
Unfocused	v_3	S_3	$max(level_p, level_y, level_r) = 3$
Very unfocused	v_4	S_4	$max(level_p, level_y, level_r) = 4$

hierarchical analysis method, the judgment matrix H is constructed as

$$H = \begin{bmatrix} 1 & 1 & 5 \\ 1 & 1 & 5 \\ 0.2 & 0.2 & 1 \end{bmatrix}. \tag{5}$$

The weight value $W = w_1, w_2, w_3 = [0.45455, 0.45455, 0.09091]$ is obtained through the square root method [14] and the largest characteristic root $\lambda_{\max}$ is 3. According to Table 5, the corresponding RI value is 0.52 [15]. The consistency test result $CR = 0 < 0.1$ is obtained from

$$CI = \frac{1}{3}(\lambda_{\max} - 3), \tag{6}$$

$$CR = \frac{CI}{RI}, \tag{7}$$

which passes the consistency test.

A single-factor fuzzy evaluation of the three Euler angles is performed separately to obtain the single-factor evaluation matrix

$$R_{1i} = \begin{bmatrix} r_{11} & \dots & r_{14} \\ \vdots & \vdots & \vdots \\ r_{41} & \dots & r_{44} \end{bmatrix} = (r_{pq})_{4\times 4}, i = 1, 2, 3. \tag{8}$$

Table 5. Consistency test RI value

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46

And r_{pq} denotes the degree of affiliation of No.p factor in the second factor sets to No.q element in the evaluation sets and satisfies

$$\sum_{q=1}^{4} r_{pq} = 1, p = 1, 2, 3, 4. \tag{9}$$

Next, the fuzzy comprehensive model $B = [B_{11}, B_{12}, B_{13}]^T$ is established according to the one-factor evaluation matrix of Euler's angle, and the calculation is shown in

$$\begin{aligned} B_{1i} &= W_1 \cdot R_{1i} \\ &= [w_{i1}, w_{i2}, w_{i3}, w_{i4}] \cdot \begin{bmatrix} r_{11} & \cdots & r_{14} \\ \vdots & \ddots & \vdots \\ r_{41} & \cdots & r_{44} \end{bmatrix}, i = 1, 2, 3, \end{aligned} \tag{10}$$

where the first layer of the factors are considered to have the same weight, i.e., $w_{i1} = w_{i2} = w_{i3} = w_{i4} = 0.25$. After obtaining the fuzzy comprehensive matrix of Euler's angle, according to the different combinations of weights, we get the fuzzy comprehensive evaluation matrix of students' concentration D, as shown in

$$D = W \bullet B = [w_1, w_2, w_3] \bullet \begin{bmatrix} B_{11} \\ B_{12} \\ B_{13} \end{bmatrix}. \tag{11}$$

Finally, according to the evaluation set, we convert students' concentration into percentage scores and set the scores of S_1-S_4 as 100, 75, 50, 25 respectively, and the concentration scores are shown in

$$Y = D \bullet S = D \bullet \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{bmatrix}. \tag{12}$$

3 Experimental Validation and Analysis of Results

This section completes the experimental validation of the overall algorithm and analyses the results of the experiments. The effectiveness of the modules of the algorithm was verified through facial key points detection experiment and head attitude angle measurement experiment. The implementation of the overall functional effectiveness of the algorithm was tested under the classroom video dataset of 35 students.

3.1 Face Key Points Detection Experiment

The main parameters for setting up Retinaface are as follows.

- Threshold = 0.8. Candidate boxes with probability values greater than this threshold are subjected to the next cropping step.
- Scales = [1.0]. It is used to detect images at different scales to improve the accuracy of detection. The default value is [1.0], which means no scaling is performed.
- Do_flip = 0. It represents no image flipping.

The face detection effect is tested in a classroom video with 35 students. The video frame rate used in the experiment is 25 frames per second. A 40-second classroom video was randomly intercepted for testing and sampled every 25 frames, and the detection rate is shown in Fig. 2.

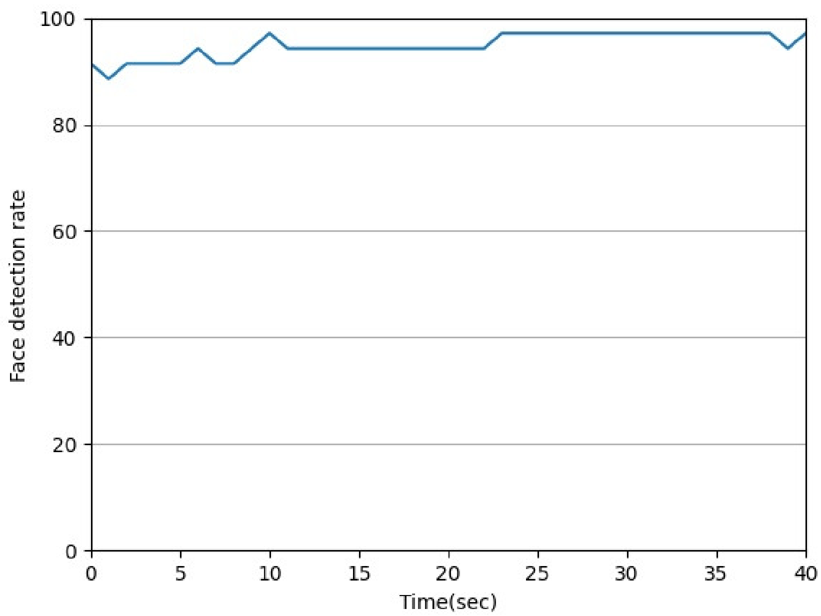

Fig. 2. Face detection rate for classroom video

The significant effects of the five face key points are shown in Fig. 3.

The visualization of face key points detection is shown in Fig. 4. In the classroom video, the average face detection rate of the 40-second video is 90.66%, and most of the undetected faces are those where the student's face has completely disappeared (e.g., lying on a table) or has been obscured. The accuracy of the

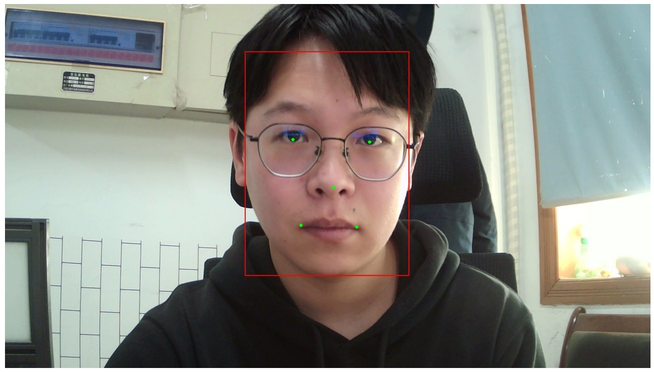

Fig. 3. The significant effects of the five face key points

Retinaface algorithm meets the demand for key points detection of faces in the classroom. Some of the cropped single-face images are shown in Fig. 5.

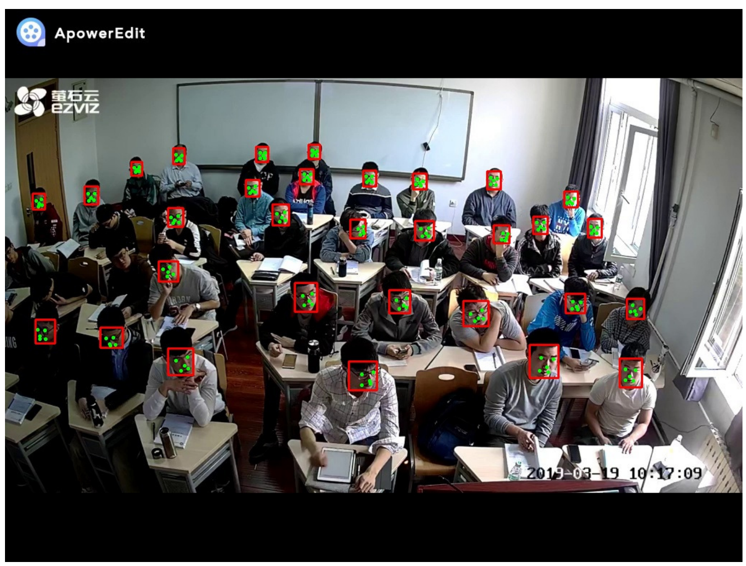

Fig. 4. The visualization of face key points detection

3.2 Head Attitude Angle Measurement Experiment

The three states of head down, head tilt to the right, and head turns to the right were tested on a single-person video dataset to visualize the effect of head

Fig. 5. Part of cropped face images

attitude angle measurement, as shown in Fig. 6. The three numbers indicate the pitch angle, yaw angle, and roll angle in turn.

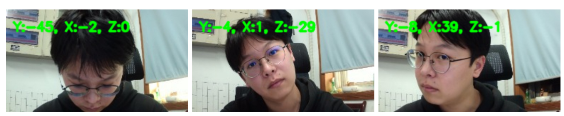

Fig. 6. Euler angles measurement results for one-person experiment

The head attitude Euler angle is calculated separately for each individual based on Retinaface cuts of individual face images. The head attitude angle measurement experiment under the classroom surveillance video dataset is shown in Fig. 7.

According to the visualisation results shown in Fig. 6 and Fig. 7, and the theory described in Eqs.(1– 4). It can be seen that the Euler angles obtained by the head attitude angle measurement module proposed in this paper basically conform to the objective situation. In the experiments of complex offline classrooms, the accuracy of head attitude angle measurement depends on the accuracy of face key points detection.

3.3 Experimental validation based on classroom surveillance video

Random 20 minutes of video data are intercepted from the classroom surveillance video of 35 students. The video includes a 5-minute class break between minutes 6 and 11, with vertical dashed lines at the time points of minutes 6 and 11. The effectiveness and validity of the algorithm's application in an offline classroom are tested on the 20-minute video. The frame rate of the experimental video is 25 frames per second. Video images are sampled every 750 frames (i.e., every 30

Fig. 7. Euler angles measurement results for classroom video experiment

seconds). After reading the video, a line graph of students' concentration scores over time is obtained, as shown in Fig 8. We analyze the algorithm's real-time performance to ensure its suitability for real-world classroom environments. We use an AMD Ryzen 9 5950X CPU for inference, and the algorithm outputs the student's concentration score at 30 FPS. This meets the requirement for a real-time classroom. Furthermore, by optimizing hardware and implementing parallel processing technology, we can further reduce inference delays and enhance the algorithm's applicability in the classroom.

As seen in Fig 8, fluctuations in students' concentration scores were detected in the first 6 minutes but remained at a high level overall. Between the 6th and 11th minutes, a class break was taken, during which concentration scores were at a lower level. At the 11th minute, when the recess ended and the class resumed, students' concentration scores increased and gradually returned to the state they were in 6 minutes before the end of the class.

Ten frames of classroom video were randomly intercepted. Fifteen reviewers were asked to rate the students' concentration in the frames, and the average of scores from reviewers was used as an objective criterion for the evaluation of students' concentration. Comparing it to the algorithm's score obtained by

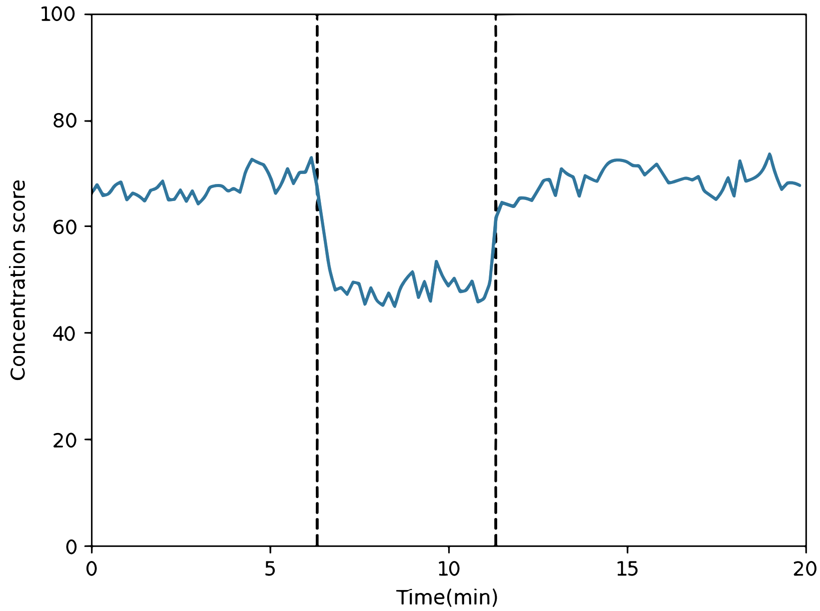

Fig. 8. Concentration results from classroom video experiment

Equation 12, the results are shown in Table 6. Comparing the expert's score with the algorithm's score, the average accuracy was 88.3%. The Pearson correlation coefficient was used as a concentration index to further verify the effectiveness of the algorithm. The formula is shown in

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad , \tag{13}$$

where the closer the absolute value is to 1, the stronger the correlation between the two variables [16].

Table 6. Comparison of Algorithm's Score

Expert's score	72.5	61.8	80.4	42.4	40.2	57.3	81.2	65.3	45.4	75.6
Algorithm's score	64	68	71	49	47	63	74	61	52	67
Accuracy	88.3%	90.0%	88.3%	84.4%	83.1%	90.1%	91.1%	93.4%	85.5%	88.6%

The Pearson correlation coefficient of 0.936 can be obtained from Equation 13, which proves that the algorithm's concentration scores have a high correlation with the artificial scores, verifying the effectiveness of the algorithm.

The algorithm is suitable for evaluating students' concentration in offline classrooms.

4 Conclusion

In order to evaluate students' concentration in offline education, an evaluation algorithm based on a fuzzy comprehensive evaluation is proposed. In the algorithm, Retinaface is employed to extract face images of students from surveillance video frames and detect 2D key points on the faces. These 2D key points are matched with 3D standard face models, enabling the derivation of Euler angles representing head attitude through a rotation matrix. Subsequently, all the Euler angles of the students of an image frame are taken as input. Data fusion is performed through a comprehensive fuzzy evaluation to obtain the student's concentration score. Once the video is analyzed, a concentration score curve over time is generated.

The average face detection rate of 90.66% in a 40-second video verifies the accuracy of face key points detection. The algorithm outputs the student's concentration scores at 30 FPS, meeting the requirement of a real-time classroom. The validity of head attitude angle measurement is confirmed in a one-person video experiment. The overall functionality of the algorithm is tested on 20 minutes of real classroom video, resulting in an average accuracy of 88.3% and a Pearson's correlation coefficient of 0.936 compared with artificial scores. The accuracy and correlation coefficient value validate the overall effectiveness of the algorithm, demonstrating that the concentration scores obtained align with the objective situation.

Compared with other concentration evaluation methods, the proposed algorithm evaluates overall student concentration using low-pixel video. Therefore, it can run in most offline classrooms with surveillance devices and help realize the automatic evaluation of education quality in the future. The effectiveness of the algorithm is further discussed in classrooms with different settings, which have different shooting angles, lighting conditions and occlusions. For different shooting angles, the standard value of the Euler Angle corresponding to the state of the front platform is different, and the standard Euler Angle can be adjusted to adapt to the concentration evaluation of the classroom with different shooting angles. Different lighting conditions and occlusions will affect the accuracy of the algorithm; too strong or too dark light, as well as too many occlusions, will reduce the accuracy of face key point detection and then affect the concentration evaluation, which is inevitable in the algorithm. In future research, considering that face detection is unable to detect the complete loss of facial information, we will try to add the criterion of action state recognition to the scoring rules. This addition will enable the algorithm to achieve a more reasonable and effective evaluation of students' concentration.

References

1. A. V. Savchenko, L. V. Savchenko, and I. Makarov, "Classifying emotions and engagement in online learning based on a single facial expression recognition neural network," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 2132–2143, 2022.
2. C. Pabba and P. Kumar, "An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition," *Expert Systems*, vol. 39, no. 1, p. e12839, 2022.
3. C. Thomas and S. Kumar, "Predicting student engagement in classrooms using facial behavioral cues," *Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education*, 2017.
4. M. Zhong, J. Zhang, Y. Lan, and Y. He, "A study on online education focus based on face detection and fuzzy comprehensive evaluation," *Journal of Computer Science*, vol. 47, no. 196-203, 2020 (in Chinese).
5. J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-shot multi-level face localisation in the wild," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5203–5212.
6. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
7. J. Zhou, "Vision-based human pose estimation in smart classroom," Master's thesis, Shanghai Jiao Tong University, 2011 (in Chinese).
8. Y. Tang, Z. Sun, and T. Tan, "A survey on head pose estimation," *Pattern Recognition and Artificial Intelligence*, vol. 27, no. 213-225, 2014 (in Chinese).
9. A. Janota, V. Šimák, D. Nemec, and J. Hrbček, "Improving the precision and speed of euler angles computation from low-cost rotation sensor data," *Sensors*, vol. 15, no. 3, pp. 7016–7039, 2015.
10. S. Yang, Y. Dai, S. Li, and K. Zhao, "An automatic analysis and evaluation system used for teaching quality in mooc environment," in *2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI)*, 2021, pp. 38–41.
11. L. Galleani and P. Tavella, "Time and the kalman filter," *IEEE Control Systems Magazine*, vol. 30, no. 2, pp. 44–65, 2010.
12. J.-F. Chen, H.-N. Hsieh, and Q. H. Do, "Evaluating teaching performance based on fuzzy ahp and comprehensive evaluation approach," *Applied Soft Computing*, vol. 28, pp. 100–108, 2015.
13. K. Peniwati, "Group decision making: Drawing out and reconciling differences," *International Journal of the Analytic Hierarchy Process*, vol. 9, 12, 2017.
14. H. Liao, Y. Qiu, X. Yang, X. Wang, and R. Ge, "A study of weight coefficient computing method based on ahp," *Mechanical Engineer*, no. 22-25, 2012 (in Chinese).
15. P. S. . S. Ya, "Research on the application of analytic hierarchy process in college teaching evaluation system," *Light Industry and Science Technology*, vol. 37, no. 167-169, 2021 (in Chinese).
16. H. Xu and Y. Deng, "Dependent evidence combination based on shearman coefficient and pearson coefficient," *IEEE Access*, vol. 6, pp. 11 634–11 640, 2018.