A deep learning based online classroom fatigue monitoring system for students

Zijie Wang[a], Liu Long[a], Yunwen Lang[a], Yuanxia Ji[b], Jie Xie[a], Shaokun Lu[*a]
[a]School of Big Data, Yunnan Agricultural University, Kunming 650201, Yunnan, China; [b]Graduate Department, Yunnan Agricultural University, Kunming 650201, Yunnan, China

ABSTRACT

To effectively monitor students' online classroom fatigue, this paper uses the improved YOLOv5s target detection model and the Dlib library to detect students' classroom fatigue. First, the improved YOLOv5s face detection model is used to detect faces instead of the detection model in Dlib, and then the detected face images are input to the official open source Dlib library, and key parts of students' mouths, eyes and heads are extracted using 68 face key point detectors. Then the student's visual localization and facial features are fused, followed by the PERCLOS algorithm to give the new metrics EAR (eye aspect ratio) and MAR (mouth aspect ratio) of the student's subject fatigue. The EAR, MAR and HPE (Human Posture Estimation) algorithms are also combined to calculate the student's eye area, mouth area and head posture parameters. Finally, according to the set thresholds, students are detected and alerted to fatigue from three indicators: blinking frequency, yawning frequency and nodding drowsiness frequency. The proposed method is an effective method for monitoring students' online classroom fatigue.

Keywords: Students, improved YOLOv5s model, Dlib library, fatigue detection

1. INTRODUCTION

The Ministry of Education of the People's Republic of China put forward the concept that classes are closed but education and learning do not stop. The purpose of this concept is to enable students to learn even under the new crown pneumonia epidemic. Due to the nature of online learning, which cannot be monitored in real time, there is often a phenomenon that students' fatigue increases gradually with the length of teaching and the information on students' fatigue cannot be fed back to teachers in time, which seriously affects students' absorption of knowledge and also leads to low teaching quality of teachers. To address this situation, it is necessary to monitor students' fatigue level in online teaching, so as to monitor students' listening status and remind them of their fatigue in time. In addition to this, accurate and timely feedback to the teacher on each student's listening status will also help the teacher to adjust the teaching style in a timely manner, thereby improving the quality of the lesson.

2. YOLOV5S AND IMPROVEMENT METHODS

2.1 YOLOv5s model

The Yolo target detection algorithm uses convolutional neural networks to regress the location of the target and get the category, enabling end-to-end real-time target detection with faster speed and generalization. Compared with the traditional detection methods in Dlib, Yolo has better robustness, smaller model size, and can better meet the real-time requirements, which makes it more suitable for applications in online network environment[1].

YOLOV5s target detection model consists of four parts. (i) Backbone network, which serves to extract the image features; (ii) Head detection head, which serves to predict the target frame and predict the class of the target; (iii) Neck layer between the backbone network and the detection head; and (iv) Prediction layer that outputs the detection results and predicts the target detection frame and label class[2].

---

[*] 3847462@qq.com

2.2 Improvement methods

2.2.1 Improvement of Data Processing. In this paper, there are two main improvements in data processing: data increment and label smoothing. Increasing the amount of data can prevent overfitting and improve the generalization ability of the model. Robustness is an important index to evaluate the stability of the system, so noise is added appropriately. The datasets in the experiment are from the COCO dataset and web crawler respectively, and the combined two parts of the dataset are expanded and enhanced using image enhancement techniques to improve the recognition accuracy. From the perspective of preventing overfitting, label smoothing is also a regularization method that can effectively prevent overfitting and can make the recognition accuracy of the system on the test set close to the recognition accuracy of the system in the real environment and improve the generalization ability of the system.

2.2.2 Improvement of Loss Function. GIoU is used as the loss function of the Yolov5s detection model, as shown in equation (1).

$$L_{GIoU} = 1 - IoU(A, B) + \frac{|C - A \cup B|}{|C|} \tag{1}$$

In the above equation, A represents prediction frame, B represents Ground Truth, and C represents closure of A and B. GIoU expands the prediction frame, hoping to make the prediction frame close to Ground Truth. However, in the subsequent calculation, it makes the convergence slow in order to allow the two boxes to overlap. To solve this problem, the DIoU loss function is used in this experiment, as in equation (2)[3].

$$L_{GIoU} = 1 - IoU(A, B) + \frac{\rho^2(A_1, B_1)}{c^2} \tag{2}$$

where A, B remain the prediction frame and Ground Truth, respectively, A1B1 denotes the centroid of the two frames, and ρ is the Euclidean distance between the two frames. In addition, c denotes the distance of the diagonal of the closed region of the two boxes. the DIoU loss is able to shrink the distance between A and B to the minimum value, which makes the two boxes coincide faster and thus speeds up the convergence[4].

3. MODEL PERFORMANCE COMPARISON

The operating system for this experiment is Windows 10, the CPU is 11th Gen Intel(R) Core (TM) i7-11800H, the GPU is GeForce RTX3060 with 8GB video memory, and Pytorch is used as the development framework. The data set is divided according to 80% training set, 10% validation set, and 10% test set, and the improved model is tested separately from the original model. The number of data loaders is 4, the training and test image sizes are set to 640*640, and some experimental hyperparameters are shown in Table 1.

The average accuracy (IOU=0.5) of the improved model is shown in Figure 1 after 250 rounds of training with the original Dlib target detection model in the same configuration. The improved model is a and the original Dlib model is b. The horizontal coordinates are the number of training rounds and the vertical coordinates are the values, both unitless. It can be observed that both models converge rapidly in the first 50 rounds, and gradually stabilize after 100 rounds until the end of training, and both models are well trained without overfitting or underfitting. And the improved model has a significant improvement in the average accuracy rate compared with the original model, which verifies the feasibility of the improved strategy.

Table 1. Hyperparameter settings.

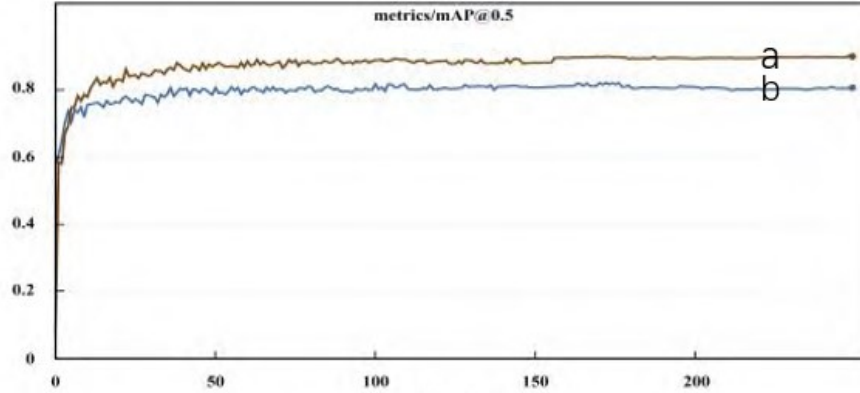| Parameters	Numerical value
Initial learning rate	0.01
Termination rate of study	0.2
Learning rate adjustment rounds	5
Number of incoming images at a time	8
Number of training rounds	250

Figure 1. mAP@0.5 curve comparison.

4. OVERVIEW OF FATIGUE DETECTION ALGORITHM

4.1 PERCLOS algorithm

PERCLOS is a recognized and valid measure of psychophysiological fatigue it is expressed as the percentage of total time that the eyes are closed above a set threshold in 30 seconds or 1 minute, it is the percentage of time that the eyes are closed above 80% of the unit time. It is commonly used in the following criteria: P70 (i=70%), P80 (i=80%), and EM (i=50%), each of which indicates tight eye closure when the pupil is blocked by the eyelid area by more than a percentage[5].
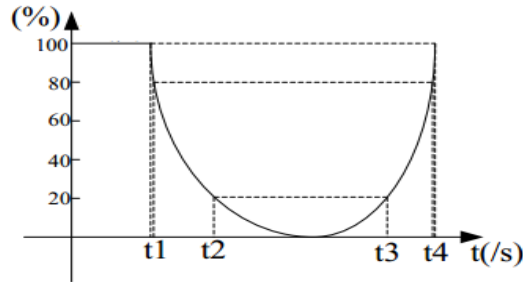

Figure 2. P80 detection schematic.

As shown in Figure 2, when the eye is closed once, the time required for the different stages is expressed in terms of t1, t2, t3 and t4, which gives the P80 standard for fatigue detection equation (3).

$$P80 = \frac{t3-t2}{t4-t1} X100\% \tag{3}$$

In video streaming images, the general formula for PERCLOS is equation (4), and as the data gets larger, the greater the proportion of the total number of frames indicating eye closure, the more severe the fatigue.

$$PERCLOS = \frac{Number\ of\ frames\ with\ eyes\ closed}{Total\ number\ of\ frames\ in\ the\ detection\ time\ period} X100\% \tag{4}$$

4.2 Fatigue characteristics determination

4.2.1 Fatigue Determination Based on Eye Features. This study is using the open source model in Dlib, the 68 feature point detection model for the face, to detect the position of the eyes. The distribution of feature points is shown in Figure 3. Of the 68 dimensional points that form the shape of the face, the left eye is the 37th-42nd dimensional point and the right eye is the 43rd-48th dimensional point. A standard P80 value can be calculated from the eye feature points.
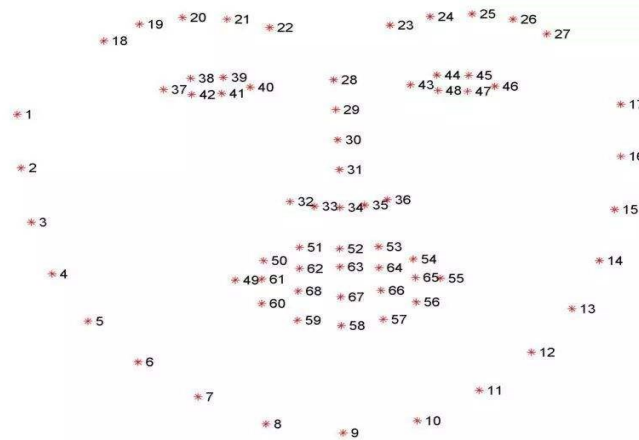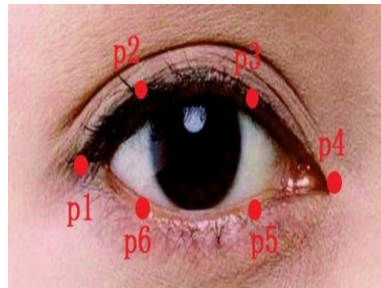
Figure 3. Distribution of 68 feature points.

Figure 4. Map of eye features.

Blinking, a rapid eye closing action, becomes more frequent when students are in a state of fatigue. In this paper, an EAR detection method, the eye aspect ratio, is applied. According to the 68 feature points that can be detected in the previous paper, let the eye position be defined as P1-P6 respectively, the distribution is shown in Figure 4, and the formula for calculating EAR is shown in equation (5).

$$EAR = \frac{\|P2-P6\|+\|P3-P5\|}{2\|P1-P4\|} \qquad (5)$$

In general, the eye feature point is constant and changes in head position and posture have no effect on it. When the eyes are closed, however, the value of the EAR drops rapidly for a moment until it approaches 0. As shown in Figure 5, a period of time when the value of the EAR is steady is judged to be a period of time when the eyes are open, and a moment when the EAR suddenly drops and then rises to the value of the previous steady state is judged to be a moment when a blink is made.
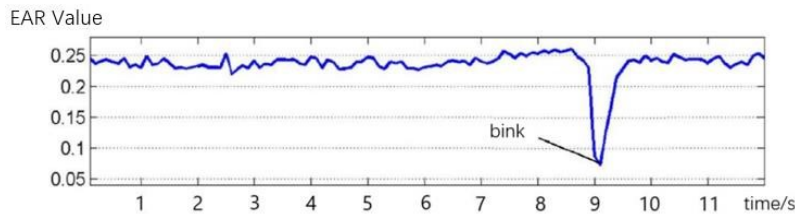

Figure 5. Blink process over a period of time.

The right eye opening and closing degree is calculated as shown in equation (6). The eye opening and closing degree is calculated from the vertical coordinates of the four characteristic points above and below the right eye and the horizontal coordinates of the two characteristic points on the left and right, with the opening and closing degree being expressed from small to large for the eye opening period and from large to small for the eye closing period. The fatigue of the person being

tested is determined by the number of eye closures and the longest period of eye closure during the testing time. In this case, in the video stream, we can use the number of frames to indicate the longest time that the subject's eyes were closed.

$$EyeOpenness = normalize\left(\frac{P42 \cdot y + P41 \cdot y - P38 \cdot y - P39 \cdot y}{P40 \cdot x - P37 \cdot x}\right) \tag{6}$$

4.2.2 Fatigue Determination Based on Yawn Frequency. Yawn belongs to a special deep breathing action, when people are in a state of fatigue, they will unconsciously produce this action, and the more frequent yawning, the more obvious it is that sleepiness. Therefore, this physiological reaction can be used to visualize the fatigue state of a person. Figure 6 shows the location of the six characteristic points of the mouth.
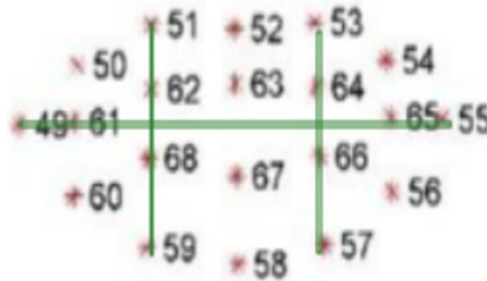

Figure 6. Map of the characteristic points of the mouth.

In this study, fatigue was determined using the MAR (Mouth Aspect Ratio), which is calculated as in equation (7), where the difference between the longitudinal coordinates of points 51, 59, 53 and 57 and the MAR value increase as the degree of mouth opening increases; conversely, the difference between the longitudinal coordinates and the MAR value decrease rapidly as the degree of mouth opening decreases.

$$MAR = \frac{\|Y51 - Y59\| - \|Y53 - Y57\|}{2\|X55 - X49\|} \tag{7}$$

4.2.3 Fatigue Determination Based on Head Posture. In this study, a HPE algorithm is introduced for determining fatigue based on head pose. The basic steps of the HPE algorithm are: detecting 2D face key points; matching 3D face models; solving the correspondence between 3D points and 2D points, and solving the head pose Euler angles based on the rotation matrix. This process needs to refer to four coordinate systems in the field of computer vision, where UVW denotes the world coordinate system, XYZ denotes the camera coordinate system, uv denotes the image center coordinate system, and xy denotes the pixel coordinate system[6]. Figure 7 shows the distribution of the four coordinate systems put together.
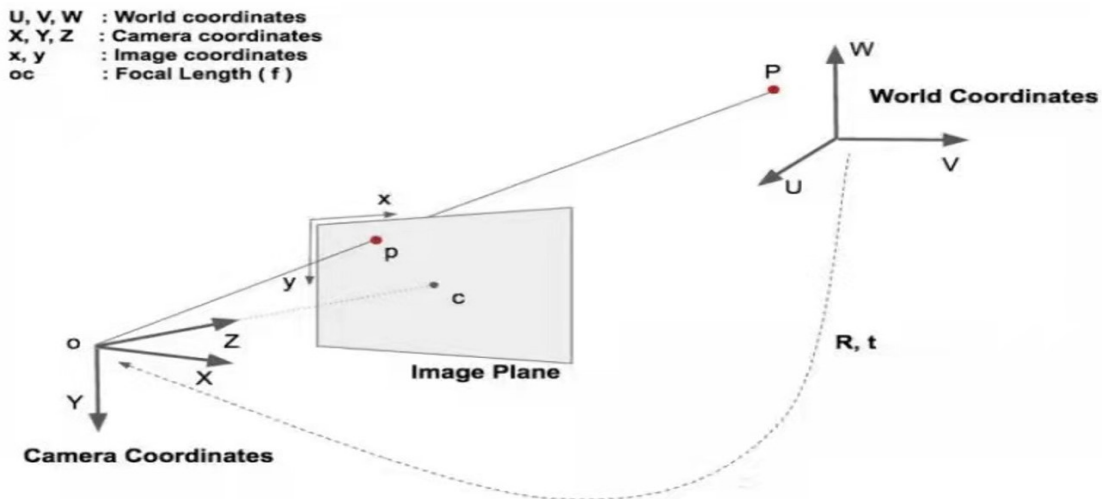

Figure 7. Distribution of the four coordinate systems.

When using the HPE algorithm to determine fatigue, there are two ways to do this:

(1) The values of Pitch (pitch angle in head rotation) and Roll (roll angle in head rotation) are used to determine whether the head is moving backwards and forwards (nodding) or tilting left and right, and the amplitude of head nod and head tilt.

(2) Using feature point 30 of the 68 feature points on the face and the distance travelled by feature point 30 during the detection time determine the magnitude and probability of nodding. The distance travelled is proportional to the magnitude of the nod, with greater distance indicating greater probability and magnitude of the nod[7].

5. SYSTEM DESIGN AND IMPLEMENTATION

5.1 System architecture

In order to better determine the level of student fatigue, we have carefully observed a large number of videos of students in catechism studies and have summarised the following fatigue-related characteristics. The camera only captures an image of the upper body of the student while the student is online in class. In the upper body, the student's eyes may or may not be on the computer screen. The head may appear to rotate in all directions. The head may be rotated in all directions and the body may be seated, tilted and submerged. There is more characteristic information about the student's face, which can be used to determine whether the student is tired by the frequency/number of blinks; the frequency/number of drowsy postures; and the frequency/number of yawns. In this experiment, students' facial fatigue characteristics were examined, formulas for determining blink frequency, eye opening, head bowing and yawning were given, and then by fusing visual localisation and facial fatigue features, fatigue indicators for student subjects given with the help of PERCLOS ideas (Figure 8).
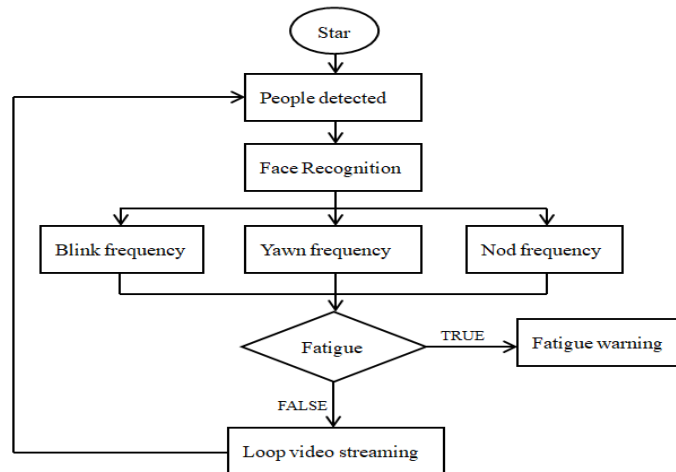
Figure 8. System implementation flow chart.

5.2 System implementation

(1) Video stream file is extracted by OpenCV. (2) Images are read from looping frames in the video stream and dimension expansion and gray scale conversion on the image are performed. (3) 68 key points are detected on a 2D human face and information are obtained on the location of the key points on the face. Then the location information will be converted into an array format and drawn. The threshold is set in advance according to the actual situation. (4) The coordinates of the left and right eyes are extracted and the EAR value is calculated respectively, and the average of the two values is taken as the final EAR value. (5) The coordinates to the mouth are extracted, the function is constructed, and the MAR value of the mouth feature is calculated. (6) The 3D face model is matched, the correspondence between 3D points and 2D points is solved, and the Euler angle of the head rotation pose is solved according to the rotation matrix. (7) The frequency/number of blinks, frequency/number of yawns and frequency/number of drowsy postures within a specified period of time are obtained, and a fatigue warning is issued when the set threshold is exceeded.

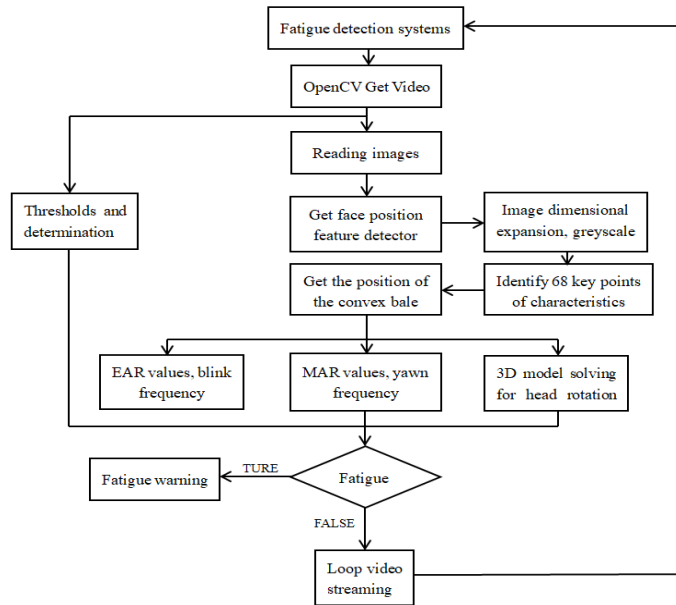The fatigue detection process is shown in Figure 9.

Fatigue detection systems

OpenCV Get Video

Reading images

Thresholds and determination

Get face position feature detector

Image dimensional expansion, greyscale

Get the position of the convex bale

Identify 68 key points of characteristics

EAR values, blink frequency

MAR values, yawn frequency

3D model solving for head rotation

TURE

Fatigue warning

Fatigue

FALSE

Loop video streaming

Figure 9. Flow chart for fatigue testing.

6. EXPERIMENTAL RESULTS AND ANALYSIS

(1) Fatigue is determined based on eye characteristics. The EAR algorithm was used to calculate the student's eye aspect ratio. In this experiment, the blink threshold was set at 0.2. When the EAR value is less than the threshold three times in a row, the number of blinks is added by one. When the number of blinks reaches 40, the system displays a fatigue warning. Only when glasses are worn is the recognition speed and accuracy lower than in the bare-eye case, in any of the other cases mentioned above the detection of fatigue is almost unaffected[8].

(2) Fatigue is determined based on yawn frequency. The MAR algorithm was used to calculate the mouth aspect ratio. In this experiment the mouth threshold was 0.5. When the MAR value is greater than the threshold for three consecutive times, the number of yawns is increased by one. When the number of yawns reaches 20, the system displays a fatigue warning. There is no effect in any of the above situations, except that the mouth is not detected when it is obscured. As can be seen from the graph below, the current mouth aspect ratio is 0.75, which exceeds the threshold of 0.5, and the yawn time is 68, so the system displays Yawning, which adds one to the number of yawns when the mouth is held closed[9].

(3) Fatigue is determined based on head posture. The HPE algorithm was used to calculate the student's head posture parameters. In this experiment, the head pitch angle threshold was 0.3. When the head pitch angle value is greater than the threshold three times in a row, the number of fishing (consecutive head nods) is increased by one. The system also displays snooze time. Fatigue is determined based on head posture and can be detected in either of these situations. As can be seen in Figure 10 below, the X value is the head pitch angle, which at this point is 17.31, greater than the threshold value of 0.3, and the snooze time is 152. When the snooze time becomes 0, i.e. when the head posture is normal, the number of fishing trips plus one[10].

Figure 10. Fatigue testing test chart.

7. SUMMARY

In this study, a deep learning-based classroom fatigue monitoring system is designed and developed in order to better practice the educational concept of "suspending classes without stopping learning" and to ensure the maximum absorption of knowledge by students and the quality of teaching by teachers. The system not only diversifies the detection indexes, but also detects fatigue normally by other indexes when one index fails, which is more fault-tolerant; it can also be used in various scenarios such as multimedia classrooms and online classrooms in the future.

ACKNOWLEDGEMENTS

REFERENCES

[1] Li, A. J., [YOLOv5 Algorithm Improvement and Its Real-Life Application], North University of China, Master's Thesis, (2021).

[2] Zhao, Y. Z. and Geng, S. L., "Face occlusion object detection algorithm based on improved Yolov5 method," Changjiang Information & Communications 34(11), 32-35 (2021).

[3] Zang, Y., [Study of IoU Loss Function in Target Detection], Anhui University of China, Master's Thesis, (2021).

[4] Huang, Z. H., Zhao, H. M. and Zhan, J., "A target tracking algorithm for Siamese network based distance intersection over union (DIOU) regression," Journal of Yangzhou University (Natural Science Edition) 24(3), 48-54 (2021).

[5] Zheng, W. C., Li, X. W. and Liu, H. Z., "Fatigue driving detection algorithm based on deep learning," Computer Engineering 46(7), 21-29 (2020).

[6] Wang, X., Zhou, X. F. and Liu, B. L., "Driver fatigue detection system based on Dlib library," Internet of Things Technologies 11(12), 26-29 (2021).

[7] Li, Y. Q., "Designed on intelligent detection system for classroom performance based on multivariate data fusion," Wireless Internet Technology 17(06), 66-68+71 (2020).

[8] Wang, Q., [Research and Implementation of Key Technology of Student Fatigue State Detection Based on Convolutional Neural Network], Central China Normal University, Master's Thesis, (2016).

[9] Chen, Y. B., Zi, Y. F. and Yang, M. Y., "Analysis of classroom teacher-student interaction based on Yolo," International Core Journal of Engineering 8(5), (2022).

[10] Ma, C. Z. and Yang, P., "Research on classroom teaching behavior analysis and evaluation system based on deep learning face recognition technology," Journal of Physics: Conference Series 1992(3), (2021).