J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 5491--5510

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

The Role of Artificial Intelligence Algorithms in English Education in Colleges and Universities and Their Teaching Strategies

Li Yang[1,✉]

[1] *Shanxian College, Heze University, Heze, Shandong, 274015, China*

ABSTRACT

The rapid development of artificial intelligence algorithms makes them play an important auxiliary role in college English education. This paper deeply analyzes the application of artificial intelligence algorithms in English education in colleges and universities, and constructs a method of analyzing students' behavior in college and university English classrooms with computer vision as the algorithmic representative, which assists teachers to understand the state of students in the whole classroom.YOLOv7 network carries out multi-target detection in the classroom and improves the network for the deficiencies in the classroom environment. The improved K-means algorithm is then introduced to improve the DeepSORT multi-target tracking algorithm. Obtain the surveillance video data in the English classroom of Q college and build the dataset by itself, and design different experiments to verify the effectiveness of this paper's algorithm respectively. Finally, the classroom behavior analysis method based on computer vision in this paper is applied to teaching practice to explore the practical application effect of the method. The results show that the improved method of this paper can significantly improve the performance of the target detection and tracking model, and the application of this paper's method to the classroom time can accurately capture the classroom state of different students, and assist teachers in formulating different teaching strategies according to different classroom stages.

Keywords: artificial intelligence, YOLOv7, DeepSORT, K-means, English education

1. Introduction

With the continuous development of science and technology, artificial intelligence technology has been widely used in various fields, and the field of education is no exception [1-2]. In the teaching of English

✉ Corresponding author.
E-mail address: 15053029358@163.com (L. Yang).
Received 22 March 2024; Revised 05 June 2024; Accepted 10 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-304

in colleges and universities, the use of artificial intelligence technology for teaching can greatly improve teaching efficiency and teaching quality [3-4].

The application of artificial intelligence technology in English teaching in colleges and universities can cover a number of aspects, such as intelligent teaching aids, personalized learning programs, natural language processing technology and so on [5-6]. Intelligent teaching aids can provide students with a more intuitive learning experience through voice recognition, image recognition and other technologies, such as automatic error correction, voice dialogues and so on [7-8]. Personalized learning programs can analyze students' learning data to develop corresponding learning plans and contents for each student, thus improving learning effects [9-10]. Natural language processing technology can help students access English learning resources and materials more conveniently, such as intelligent translation and language understanding [11-12]. Of course, the integration and development of AI and university English education also faces some challenges and problems, one of which is that the promotion and application of the technology requires a large amount of investment, and systematic technical research and development and practical exploration for English education scenarios are needed, and the other is that the privacy and security issues also need to be emphasized [13-16]. Only by making full use of the advantages of AI technology and combining the advantages of traditional education can we realize the integration and development of AI and university English education and meet the new challenges of future education [17-19].

Literature [20] and literature [21] examined the application and role of Artificial Intelligence (AI) in English language teaching and showed that AI provides a positive learning environment for English education, which can effectively improve the students' practical skills such as writing and optimize the effect of English language teaching. Literature [22] examined the application value of AI in language learning, and analyzed the application of AI in English teaching in colleges and universities from the aspects of listening, speaking, writing, translating, and intelligent management, and elaborated on the innovative strategies of AI technology in university English teaching. Literature [23] discusses the application strategies of AI in English teaching based on literature review, describes the AI strategies and their application in English teaching, including the effectiveness of the application, the practical use, and the requirements for using in English teaching, and verifies the effectiveness of AI in English teaching. Literature [24] outlines the impact of AI on the way English is taught in higher education and the digital and personalized learning provided, and examines the strategies for integrating Al into English language teaching in higher education, concluding that an augmented whale supertuned artificial neural network optimizes the teaching of English at the university level, especially the improvement of spoken language. Literature [25] explores the application of AI technology in the teaching of English translation, reveals that AI technology may have a significant impact on the development of future translators' competence, and proposes a teaching concept for an online education course on "Synchronous and Asynchronous Translation in Digital Environments". Literature [26] constructed an English learning and teaching system based on the combination of semantic Web technology and AI technology, and combined with the needs of English teaching, constructed the functional modules of the system, analyzed the performance of the system, and indicated that the system can meet the actual needs of English independent learning and English teaching. Literature [27] introduces the multimedia and Internet of Things-assisted intelligent English teaching system, which aims to improve the stability of the system and enhance the personalized experience of students, and verifies that the system can find reliable research ideas for intelligent teaching in colleges and universities, and can effectively improve students' learning performance and narrow the achievement gap.

On the basis of analyzing the insufficiency of traditional English classroom behavior analysis in colleges and universities, this paper selects computer vision technology as a representative of artificial intelligence algorithms, and places it into the module of classroom behavior analysis in college and university English education to explore the auxiliary role of artificial intelligence algorithms in college and university English education. Then, according to the actual needs of the research in this paper, the loss function of YOLOv7 network is improved and the attention mechanism is introduced into the network, which transforms global pooling into a pair of one-dimensional feature encoding and aggregates feature information from the two directions of height and width, respectively. The small target detection head is added to the original YOLOv7 to improve the model's sensitivity to target detection. The DeepSORT algorithm is used for multi-target tracking of learning behaviors in the English classroom, and the improved K-means algorithm is introduced to deal with its bounding box for the slow target tracking speed of the algorithm. At the same time, matching variables are defined to solve the problem of the algorithm's poor performance in low-light environments. The video data of English classes in eight types of classrooms in the second half of the 2023 semester in University Q are collected, and the student behaviors in the data are defined and labeled with reference to the S-T behavioral analysis method, which is divided to form the research dataset of this paper. Experiments are used to verify the superior performance of the improved YOLOv7 algorithm and the improved DeepSORT algorithm of this paper in the self-constructed dataset, respectively. The student behavior analysis method of this paper is applied to actual English education in colleges and universities to boost the innovation and development of English education in colleges and universities.

2. Architectural design

Classroom learning behavior analysis is an important part of teaching. By observing, recording and analyzing students' learning behaviors, teachers can adjust teaching strategies in a targeted way. Traditional English education in colleges and universities focuses on the presentation of teaching content, and behavior analysis relies on teachers' subjective observation and empirical judgment, which has problems such as low efficiency and strong subjectivity. Computer vision is an important branch in the field of artificial intelligence, which uses algorithms to analyze and understand image and video data to perceive and understand the real world. Applying computer vision technology to classroom learning behavior analysis can capture, track and analyze students' learning behavior [28], so as to objectively assess students' learning status. This study intends to place the computer vision-based learning analysis module in college English education, fully demonstrating the auxiliary role and function of artificial intelligence algorithms in college English education, and promoting the improvement of teaching quality of college English education.

This study designs a classroom learning behavior analysis system based on computer vision, and its architecture follows the design concept of “front-end sensing-edge computing-cloud service”, with multi-layer decoupling, and the specific architecture is shown in Figure 1. The data acquisition layer is deployed at the teaching site, and the collected data is uploaded to the edge computing node for preprocessing, and the preprocessed data is transmitted to the cloud and the behavior analysis layer in real time through the MQTT protocol, and with the support of the cloud service, the data processing layer performs cleaning, synchronization, and fusion. Student characteristics are extracted to form behavioral description data, which are aggregated into the behavioral analysis layer to construct a spatio-temporal representation model of student behavior.

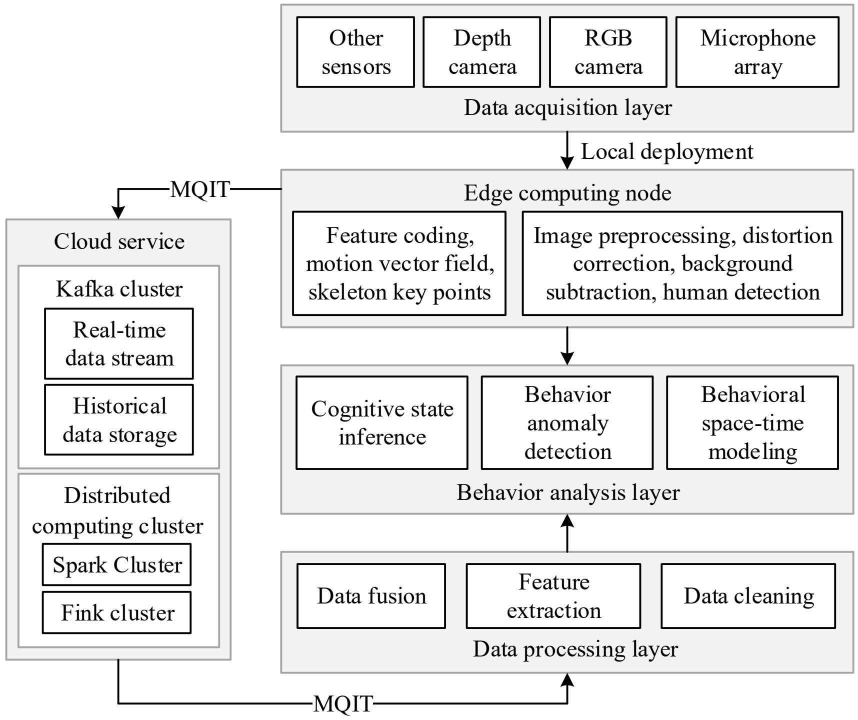


Fig. 1. Architecture of learning behavior analysis based on computer vision

3. Key methodologies

In this paper, we utilize RGB cameras at the front and back of the classroom and a panoramic camera to capture video data from the students' classroom. The captured video streams are decoded and frame extracted, then detected and tracked using the YOLOv7 target detection algorithm and the DeepSORT target tracking algorithm identity, location, posture and other information, and structured and stored as frame-level metadata records. In order to fully illustrate the student classroom behavior analysis techniques in this paper and show how AI algorithms can play a supporting role in English education in colleges and universities, this chapter will focus on the key methods in computer vision-based student classroom behavior analysis.

3.1. YOLOv7 target detection algorithm

Since this paper analyzes student behavior based on video data captured by cameras in the classroom, target detection prior to analysis is critical. When performing target detection in a classroom environment, the algorithm faces a series of challenges such as complex background noise, mutual occlusion between targets, and small target detection, which together increase the difficulty of implementing multi-target detection in complex classroom scenarios. This section provides a detailed description of the YOLOv7 network and utilizes YOLOv7 as the base algorithm for improvement studies.

3.1.1. YOLOv7 network. The network structure of YOLOv7, one of the leading target detection algorithms, can be divided into four core parts: the Input module, the Backbone network, the Neck, and the Predictive Head [29]. Firstly, the Input module scales the image to a size size of 640×640 and inputs it into the network, the Backbone module is responsible for extracting the features from the image, the Head module fuses and splices the features and outputs three feature maps with different size sizes, and finally the three prediction heads in the Detect module removes the prediction frames that overlap too much by means of the NMS, to get the final position as well as the category prediction results. The network structure is shown in Figure 2.

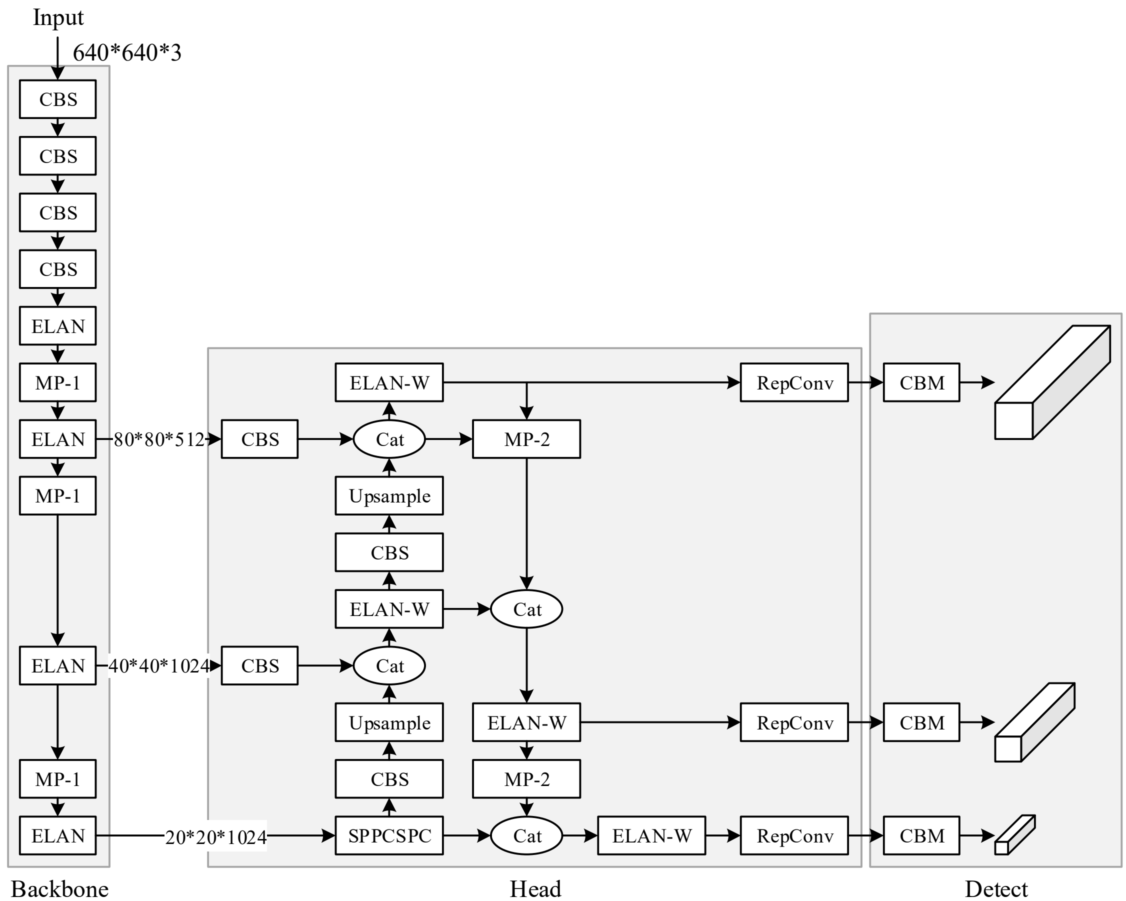


Fig. 2. Network structure of YOLOv7

Where the feature extraction unit of the network consists of CBS module, MP-1 module and ELAN module.

The CBS module consists of a convolutional layer, a BN layer and a Silu activation function, the convolutional layer in turn contains different convolutional kernels and step sizes, a 1×1 convolution is used to change the number of channels, a 3×3 convolution with a step size of 1 is used for feature extraction, and a 3×3 convolution with a step size of 2 is used to achieve downsampling.

MP-1 consists of two branches, the first branch uses maximum pooling to achieve downsampling, then a 1×1 convolution to change the number of channels, the second branch passes through a 1×1 convolution, a 3×3 convolution with a step size of 2 in turn, and finally the features extracted from the two branches are fused to achieve downsampling without feature loss.

The ELAN module is an efficient network structure that enhances the feature learning ability of the network by controlling the shortest and longest gradient paths, while the ELAN-W module adds two branches to ELAN for feature map fusion, and the number of channels in the outputs are all twice the number of inputs.

3.1.2. Improvements to the YOLOv7 network. Due to the complex background of the classroom, the targets are far away from the camera, the targets themselves are small or the image resolution is high, and the targets occlude each other. To address the problem that small targets are easy to be interfered by the background, have insufficient information or are too small in size to be detected accurately, the baseline YOLOv7 model is improved in this section. The improved detection model contains three main parts: the pre-training part, the training part and the testing. In order to reduce the complexity of target detection model training, the weights obtained by pre-training the YOLOv7 model on the dataset are used, and this operation can share some of the backbone network layer weights with the pre-training model, which makes it less difficult to train the subsequent target detection network and faster to converge. Through the self-made target training set, the loss function, attention mechanism, and small target detection layer are improved for YOLOv7 target detection network, which makes the model purposefully learn the features of the targets in the classroom scene and obtain multi-target detection results.

1) Improvement of loss function WIOU. The loss function of YOLOv7 consists of three parts: cls loss, obj loss, box loss, which are category loss, confidence loss of box, and position loss of box. A good loss function for bounding box regression can significantly improve the performance of the model. The traditional intersection and integration ratio (IOU) calculation method only considers the overlap region between the predicted box and the real box, and does not deal with the problem of vanishing of the back-propagation gradient in the case of no overlap of bounding boxes, which leads to the width of the no-overlap region can not be effectively updated in the training process. In contrast, GIOU calculates the minimum outer rectangle for each pair of predicted and real boxes, which improves the convergence speed of the model to a certain extent but still has limitations, while DIOU considers the overlap area and the distance from the centroid but ignores the effect of the aspect ratio. The CIOU function used by YOLOv7 adds a penalty term for the bounding box aspect ratio on the basis of DIOU, which further stabilizes the regression effect of the target box. However, geometric factors such as aspect ratio will aggravate the penalty for low-quality samples, and the penalty for geometric factors should be weakened when the anchor frame overlaps well with the target frame, and fewer interventions in the training will lead to a better generalization ability of the model, on which WIOUv1 is constructed, and WIOUv2 adds monotonicity coefficients based on v1, with the formulas shown below:

$$L_{WIOUv1} = R_{WIOU} L_{IOU} \tag{1}$$

$$R_{WIOU} = \exp\left(\frac{\left(x - x_{gt}\right)^2 + \left(y - y_{gt}\right)^2}{\left(W_g{}^2 + H_g{}^2\right)^*} \right) \tag{2}$$

$$L_{IOU} = 1 - IOU = 1 - \frac{W_i H_i}{S_U} \tag{3}$$

where x, y, x_{gt}, y_{gt} represents the positions of the predicted and real frames, respectively, W_g, H_g denotes the width and height of the minimum enclosing frame, and W_g and H_g are separated from the computational graph in order to prevent WIOU from generating gradients that hinder convergence. W_i, H_i is the width and height of the overlapping region, respectively, and S_U is the area of the concatenation. Then:

$$L_{WIOUv2} = L_{IOU}^{r'} L_{WIOUv1} \tag{4}$$

where $L_{IoU}^{r'}$ is represented by the ratio of the current loss L_{IOU}^{*} and the running mean $\overline{L}_{IOU}$ with momentum m. Since WIOUv2 with monotonic FM characteristics is affected by low-quality samples, it is easy to lead to poor prediction results. Therefore, this paper proposes a WIOUv3 loss function based on the improvement of WIOUv1, which introduces the dynamic nonmonotonic coefficients, as shown in the following equation:

$$L_{WIOUv3} = r L_{WIOUv1} = \frac{\beta}{\delta \alpha^{\beta - \delta}} L_{WIOUv1} \tag{5}$$

where r is the gradient gain, β is the outlier to describe the quality of the anchor frames, and α, δ takes the values of 1.9 and 3, respectively.The WIOUv3 loss utilizes β introduces a nonmonotonic focusing coefficient, which assigns small gradient gains to high-quality anchor frames with low outliers to ensure that the bounding box regression process focuses more on average-quality anchor frames. In addition, small gradient gains are assigned to low-quality anchor frames with higher outliers, effectively weakening the effect of low-quality samples on the loss of bounding box regression.

2) Embedded Attention Mechanism CA. Attention mechanism has been widely used in the field of computer vision, especially in the task of target detection, which effectively enhances the performance of the model.The core idea of CA attention mechanism is to dynamically adjust the weight of each channel according to the importance of each channel, so that the network pays more attention to the information that is beneficial to the task. To this end, this paper introduces the attention mechanism CA, the structure of which is shown in Fig. 3.The CA attention mechanism transforms global pooling into a pair of one-dimensional feature encoding, which aggregates feature information from the two directions of width and height, respectively.

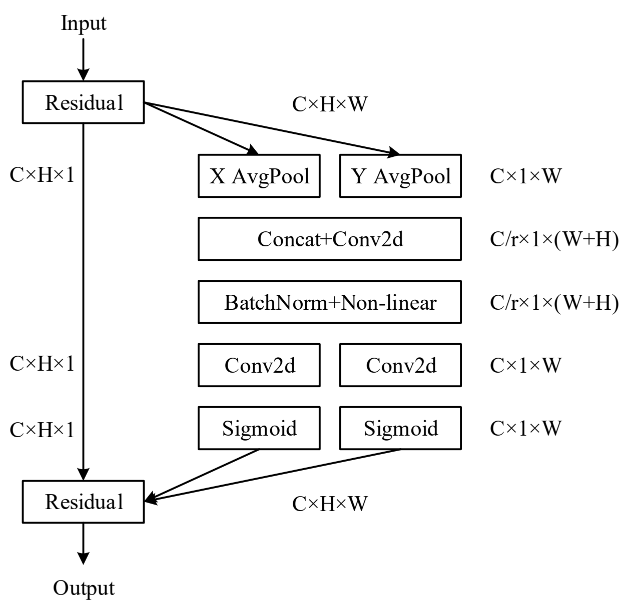


Fig. 3. CA structure

3) Addition of small target detection layer. In this paper, the small target detection head is added to the original model, which produces a feature map of 160×160 size, thus improving the model's detection sensitivity to tiny targets. The improved model initially extracts features from the backbone network and embeds the CA attention mechanism in the Head network structure, fuses the extracted shallow features with the contextual information refined by the CA attention mechanism through the use of the Concat operation, and adds a fourth detector head dedicated to small target detection. Although this improvement slightly increases the computational effort of the model, it significantly enhances the performance of YOLOv7 on small-target detection by capturing richer small-target feature information and effectively reduces the rate of misdetection and underdetection of student behaviors at different scales.

3.2. Improved DeepSORT multi-target tracking algorithm

3.2.1. DeepSORT multi-target tracking algorithm. After completing the multi-objective detection of students in the classroom, it is necessary to continuously track students' behaviors in the classroom, so as to obtain the data of students' classroom behaviors and grasp the overall performance of students in the English classroom. Therefore, this paper adopts the DeepSORT multi-objective tracking algorithm to track students' classroom behavior in the classroom scenario.

DeepSORT algorithm realizes its function by combining Kalman filtering and data matching between frames [30], and the steps of algorithm implementation are as follows:

1) Based on the detection results of the first frame, generate the corresponding trajectories and initialize the Kalman filter for these trajectories. Set the motion parameters and predict the corresponding frames using the Kalman filter.

2) Perform the matching process by comparing the intersection and concurrency ratio IoU between the target detection in the current frame and the target predicted by the trajectory in the previous frame, and develop a cost matrix accordingly.

3) Input the cost matrix from step 2) into the Hungarian algorithm to realize linear matching.

4) Repeat steps 2) to 3) until the end of the video frame.

5) After the video frames have been processed, the confirmed trajectories and unrecognized trajectories are predicted using the Kalman filter to determine their corresponding targets. Then the prediction frame of the confirmed trajectory is cascaded with the detection result using appearance features and motion information to match with the predicted bounding box of the object, and the matching result is inputted into the Hungarian algorithm so as to realize the matching between the detection frame and the prediction frame.

6) Step 5) is performed repeatedly until the end of the video frame.

3.2.2. Improvements to the DeepSORT algorithm. Since DeepSORT is characterized by slow target tracking, this paper uses the improved K-Means nearest neighbor algorithm to process the bounding box of the DeepSORT algorithm.

The original K-Means algorithm uses the Euclidean distance as the similarity measure criterion, but the clustering results obtained from the Euclidean distance are easily interfered by the large bounding boxes detected. The improved K-Means algorithm re-generates the bounding box automatically and transforms the Euclidean distance metric formula of the traditional K-Means algorithm into:

$$d(b,c)=1-I(b,c) \tag{6}$$

where: $d(b,c)$ denotes the distance between a certain bounding box b and the clustering center c. $I(b,c)$ denotes the intersection ratio between the bounding box and the clustering center.

The WIOU values of anchor frames and true bounding boxes obtained by the improved K-Means nearest neighbor clustering method are improved.The steps for solving the WIOU values are as follows:

1) Cluster the raw data containing the location and category information of the target true bounding box $(x_j, y_j, w_j, h_j), j\in\{1.2.\cdots,N\}$, where (x_j, y_j) is the coordinates of the center of the target bounding box, w_j, h_j is the width and height of the target bounding box, and N is the number of the target bounding box.

2) Initialize K clustering centers $(W_i, H_i), i\in\{1,2,\cdots,K\}$, where W_i, H_i is the width and height of the clustering centers.

3) Find the distance from the real target bounding box to each cluster center using equation (3), and then divide it into the cluster center with the closest distance to it. The distance between two bounding boxes is calculated as:

$$D_1 = 1-I((x_j, y_j, w_j, h_j),(x_j, y_j, W_i, H_i)) \tag{7}$$

$$j\in\{1,2,\cdots,N\}, i\in\{1,2,\cdots,K\} \tag{8}$$

Where: D_1 denotes the distance between two bounding boxes.

4) Recalculate the clustering center of each class cluster. The width and height of the updated ind cluster center are:

$$W_i^{'} = 1/(N_i \sum w_i), H_i^{'} = 1/(N_i \sum h_i) \tag{9}$$

Where: $W_i^{'}$ is the width of the ind cluster center after updating. N_i is the number of true target bounding boxes in the ith cluster. $H_i^{'}$ is the height of the ith cluster center after update. w_i is the width assigned to the ith cluster center. h_i is the height of all bounding boxes assigned to the ith cluster center.

5) End when the number of iterations has been reached or there is no significant change in the clustering center.

To address the problem that the multi-target tracking algorithm performs poorly in low-light environments, the DeepSPRT algorithm improves the data correlation mechanism to improve the performance under low-light conditions. To this end, a matching variable $t(m)$ can be defined for evaluating the match between the currently detected bounding box m and the reference bounding box. The matching variable $t(m)$ is calculated as:

$$t(m) = \begin{cases} 0 & \frac{S_A \cap S_B}{S_B} < \frac{1}{5} \\ \min\left\{1, 3 \times \frac{S_A \cap S_B}{S_B}\right\} & \frac{S_A \cap S_B}{S_B} \geq \frac{1}{5} \end{cases} \tag{10}$$

Where: S_A, S_B represents the area of the two detected bounding boxes for the target and the occluder, respectively. m is the current detected bounding box number.

4. Experimental analysis

In this section, the data collected from English classrooms in colleges and universities are input into the improved YOLOv7 algorithm and the improved DeepSORT algorithm proposed in this paper for experiments, so as to verify the effectiveness of the two proposed methods. Finally, the algorithms in this paper are used to conduct an example analysis of students' classroom behavior to explore the application effect of the student classroom behavior analysis method constructed based on computer vision in this paper.

4.1. Data set construction

The dataset is the most important foundation for algorithm design and validation, and a high-quality dataset has a very important impact on the training results of the model. Constructing a high-quality dataset is necessary to obtain a high-quality student behavior analysis model, which can provide support for student behavior analysis. However, research in this area is not yet mature and there is a lack of publicly available student behavior databases for research purposes. Therefore, the first priority of this study is to construct a student behavior dataset of real classroom scenarios.

In order to collect as comprehensive classroom contextual information as possible, a total of eight types of classroom videos of English classes in the second semester of the year 2023 (staircase classroom, classroom in the public education building, classroom in the arts building, classroom in the arts building, electromechanical building, food and chemical building, national heavy building, and light and chemical building) from the University of Q, with a size of close to 150 GB, were collected for this study. in the data cleaning session, the videos of these classroom videos that had students dismissing from the classroom, had a small number of students, or had students taking exams in the classroom resulting in a single action were filtered. exams resulting in a single action were screened out. In the image interception session, given that students' learning behaviors change very little in a short period of time, in order to reduce the redundancy of data and ensure the diversity, richness and validity of the dataset, this study adopts the Python programming technique to intercept a static image from the video every one minute to generate a picture in jpg format. After this operation, a total of 7,538 students' English classroom lesson image data were collected and organized.

Drawing on S-T behavioral analysis and counting the common behaviors of students in the data set, five types of student classroom behaviors were finally identified, namely, looking up to listen to the lecture, standing up to answer a question, playing with the cell phone with the head down, looking to the right and left, and reading with the head down, which were recorded as B1-B5. Based on the definition of student behaviors, the collected images were labeled with data and the number of labeled student behaviors of each type was counted. The files and images with labeled data are assigned to the training set, validation set, and test set folders according to the ratio of 6:2:2, and ensure that the images in the training set and test set come from different classrooms and different time periods, so as to ensure the independence of the data distribution and evaluate the generalization performance of the algorithm more effectively. The divided dataset is shown in Table 1. At this point, the self-constructed student behavior dataset needed for the study of student behavior analysis methods for English classes in colleges and universities is completed, and all video files, images, and annotation information will be used for subsequent research use.

Table 1. Data set partitioning results

Data set	Picture	Student behavior				
		B1	B2	B3	B4	B5
Total	7538	65472	26748	8347	9836	20834
Training set	4523	39445	16049	5008	5902	12500
Verification set	1508	13097	5350	1669	1967	4167
Test set	1507	12993	5349	1670	1967	4167

4.2. *Indicators for experimental evaluation*

4.2.1. Metrics for Evaluating Target Detection Algorithms. Evaluation metrics for target detection algorithms can quantitatively assess the training results of the algorithms, thus effectively reflecting the performance of the current algorithms and models in specific detection tasks. In this paper, the evaluation metrics selected for improving the YOLOv7 target detection algorithm are not only the commonly used F1-score, but also the mAP value to evaluate the performance of the algorithm.

The mAP is a comprehensive performance metric to evaluate the overall effectiveness of the model on different categories, including classification accuracy and the detector's ability to localize the target. mAP values are usually between 0 and 1, with higher values indicating better overall performance of the model. mAP is calculated as shown in the following equation:

$$AP = \int_0^1 P(r)dr \tag{11}$$

$$mAP = \frac{\sum_n AP}{n} \tag{12}$$

4.2.2. Evaluation metrics for multi-target tracking algorithms. In this paper, the improved DeepSORT algorithm is evaluated using Multi-Object Tracking Accuracy (MOTA), Multi-Object Tracking Precision (MOTP), and Identity ID Switching (IDSW).

IDSW indicates the total number of identity ID switches throughout the trace. MOTA is used to measure the accuracy of multi-target tracking, which is closely related to the cases of following error, following loss and ID switching, and the higher value represents the better tracking accuracy, which is calculated as shown in Equation (13), where FN denotes the number of incorrect detections per frame, FP denotes the number of undetected targets per frame, IDS denotes the number of ID switches occurring in each frame, and GT denotes the number of real objects in each frame:

$$MOTA = 1 - \frac{\sum(FN + FP + IDS)}{\sum(GT)} \tag{13}$$

MOTP is a measure of the degree of positional matching between the target trajectory predicted by the tracking algorithm and the real trajectory of the target, and the larger its value, the higher the accuracy of the tracking algorithm, which is calculated as shown in Equation (14), where t represents the tnd frame during the tracking process, $d_{t,n}$ represents the distance between the nth predicted trajectory and the real trajectory matching pair in the tth frame, and c_t represents the distance between the predicted object trajectory and the real object trajectory in the tth frame number of successful matches:

$$MOTP = \frac{\sum_{t,n} d_{t,n}}{\sum_t c_t} \tag{14}$$

4.3. Algorithm Performance Analysis

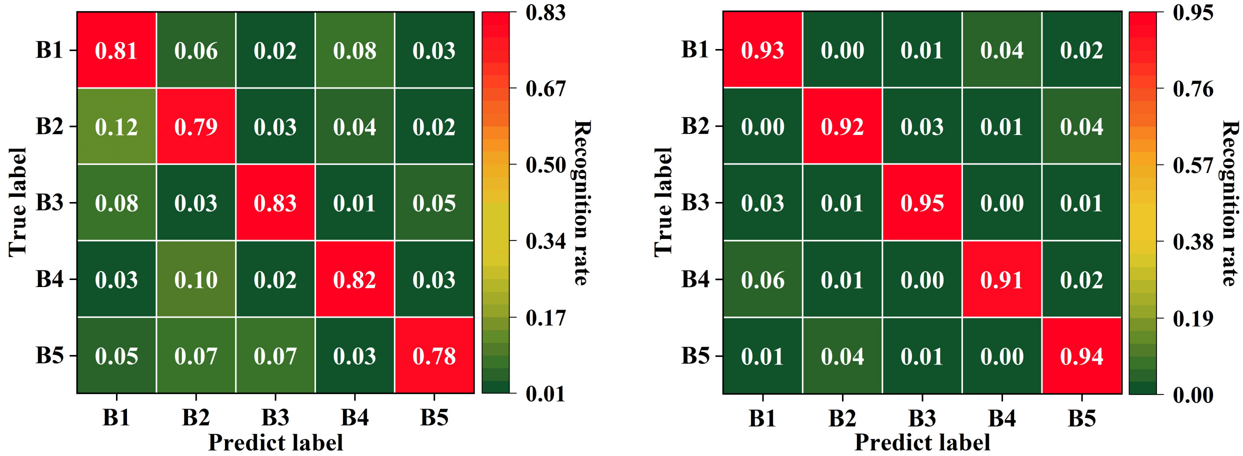
4.3.1. Performance Analysis of Improved YOLOv7 Algorithm. In this subsection, the improved YOLOv7 algorithm will be trained based on the self-constructed student behavior dataset study by choosing the most appropriate activation function, so that the improved YOLOv7 algorithm has a better target detection algorithm on the self-constructed dataset of this study.

In order to verify that the improved YOLOv7 algorithm proposed in this study has a better recognition rate than the original YOLOv7 algorithm for each type of student behavior, the training results of the two models, the original YOLOv7 and the improved YOLOv7, on the self-constructed dataset are normalized, and the visualized confusion matrix is generated using the normalization results. The vertical axis in the confusion matrix is the true labels and the horizontal axis is the predicted labels, each row represents the labeled true categories and each column represents the model-predicted categories, and the five categories of behaviors labeled in the dataset are looking up to listen to the lecture (B1), standing up to answer the question (B2), playing with the cell phone with the head down (B3), looking right and left (B4), and reading a book with the head down (B5). The confusion matrices of the original YOLOv7 and the improved YOLOv7 model are shown in Fig. 4(a) and (b), respectively. As can be seen from the figures, the improved YOLOv7 algorithm proposed in this paper improves the recognition rate compared with the original YOLOv7 algorithm in recognizing all kinds of student behaviors, especially in the case of standing up and answering a question (B2) and looking down at a book (B5), where the recognition rates are 92% and 94%, respectively, which are 13% and 16% higher than the original algorithm. Careful molecular data can also be found, although the improved algorithm has significantly improved the recognition rate of looking left and right (B4), but whether the original algorithm or the improved algorithm, the recognition rate of looking left and right is relatively low, this phenomenon may be due to the following two reasons:

1) Students looking left and right may be misrecognized as looking up to listen to the lecture when the angle of turning their heads is not obvious.

2) When constructing the dataset, the small percentage of data for looking left and right resulted in training results that were not as good as other categories.

Overall, the improved YOLOv7 algorithm achieves a recognition rate of more than 90% for all five categories of student behaviors in the dataset, which can effectively identify various types of student behaviors in the English classroom, and the recognition effect can meet the requirements of the next analysis of student behaviors in this study.



(a) YOLOv7 (b) Improved YOLOv7

Fig. 4. Confusion matrix of YOLOv7 and improved YOLOv7

In order to comprehensively verify the effectiveness and superiority of the improved YOLOv7 algorithm proposed in this study, the improved YOLOv7 algorithm proposed in this study is compared with the current mainstream target detection algorithms on the self-constructed dataset, and the results are shown in Figure 5. From the data in the figure, it can be seen that the mAP values of the YOLOv7 algorithm proposed in this study are 21.30%, 14.51%, and 11.12% higher than those of the SSD, YOLOv3, and YOLOv7 algorithms, respectively, which fully proves that the improved YOLOv7 algorithm's recognition effect on the self-constructed dataset of this research is superior to that of other mainstream target detection algorithms.

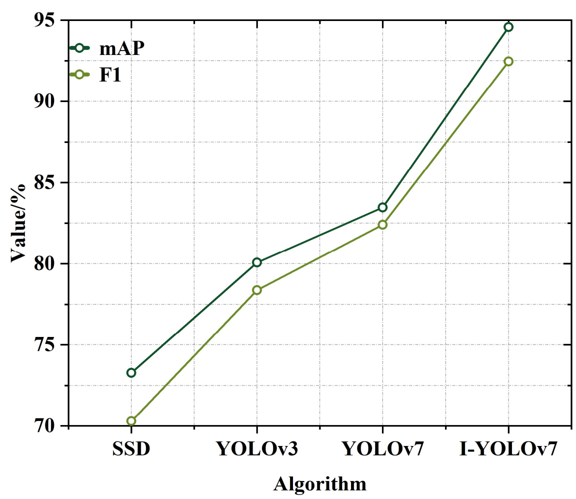


Fig. 5. Comparison of improved YOLOv7 and other target detection algorithms

4.3.2. Performance Analysis of Improved DeepSORT Algorithm. In order to verify the effectiveness of the improved DeepSORT tracking algorithm proposed in this paper for multi-target tracking of students, the improved DeepSORT tracking algorithm, the original DeepSORT tracking algorithm are experimentally compared and analyzed based on the self-constructed dataset in this paper, and the experimental results are shown in Table 2, where Model 1 represents the original DeepSORT model and Model 2 represents the improved DeepSORT model in this paper. The improved DeepSORT model. The data are fed into the tracking algorithm for experimental testing, and the performance index results of the two models for 10 experiments are obtained and the mean value is taken. By comparing the performance results of the two models under different experimental times and the mean values of the performance index results of 10 experiments, it can be seen that after improving the original DeepSORT multi-target tracking algorithm, the MOTA and MOTP of the model have been significantly improved. The mean values of MOTA and MOTP of the improved DeepSORT algorithm under 10 experiments are improved by 10.11% and 3.09%, respectively, compared with the original DeepSORT algorithm. In addition, the number of ID switches of the improved DeepSORT algorithm is reduced by 72 times on average compared to the original DeepSORT algorithm. The above experimental results fully demonstrate the effectiveness of the improvement of the original DeepSORT algorithm in this paper.

Table 2. The tracking results before and after the improvement of DeepSORT

Experimental frequency	Model 1			Model 2		
	MOTA/%	MOTP/%	IDSW	MOTA/%	MOTP/%	IDSW
1	69.75	67.89	195	79.14	72.83	124
2	69.57	69.33	194	81.49	70.35	125
3	70.56	68.89	197	79.54	69.14	122
4	70.75	69.15	195	80.08	70.57	124
5	71.69	67.32	195	81.79	72.91	124
6	70.58	67.52	197	80.21	69.55	125
7	70.04	68.98	193	81.52	70.68	123
8	69.91	66.75	195	80.78	73.61	123
9	71.34	67.65	198	80.03	73.58	123
10	69.51	70.64	196	80.21	71.74	125
Mean	70.37	68.41	196	80.48	71.50	124

In order to further verify the effectiveness of the improved tracking algorithm for multi-target tracking of students, the improved algorithm of this paper is compared and analyzed experimentally with the SORT, DeepSORT, and ByteTrack tracking algorithms on the self-constructed dataset of this paper, and the results are shown in Fig. 6. As can be seen from the figure, the MOTA, MOTP, and IDSW of this paper's improved algorithm are better than the comparison model. Compared with the SORT algorithm, the MOTA and MOTP of the improved algorithm are improved by 18.90% and 5.18%, respectively, and the number of ID switching is reduced by 143 times. Compared to the ByteTrack algorithm, the improved algorithm also improves, with MOTA and MOTP improved by 6.12% and 2.03%, respectively, and the number of student target ID switches reduced by 30 times. The comparative experimental results verify the superiority of the improved DeepSORT algorithm proposed in this paper for student multi-target tracking in classroom scenarios.

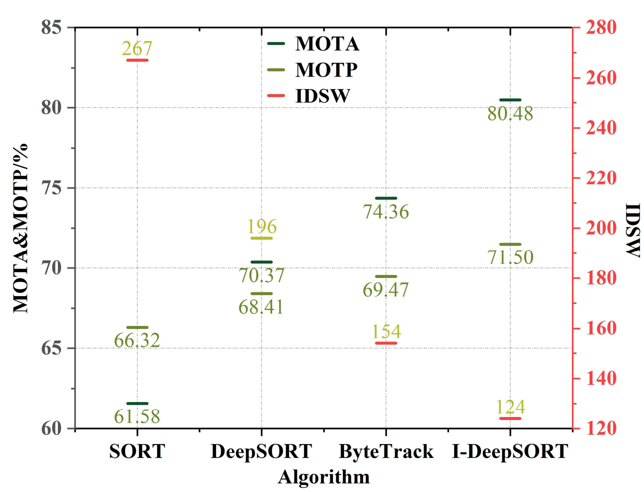


Fig. 6. Model experiment comparison

4.4. Analysis of Student Behavior in the Classroom

The previous section demonstrated the effectiveness of the improved multi-target detection and multi-target tracking algorithms proposed in this paper in classroom scenario applications through comparative experiments. In this section, these two computer vision techniques are applied to real-world scenarios to investigate the effectiveness of this paper's method for analyzing students' English classroom behavior in practice. The study also selects eight types of classrooms of English classrooms in Q University, and unlike the self-constructed dataset, the experiments in this section apply the method of this paper to the first half of the semester of 2024 to analyze the students' behaviors in the English classroom.

In order to validate the accuracy and objectivity of the method, five types of student behaviors are classified in this section as positive and negative behaviors, where positive behaviors contain three categories of looking up to listen to lectures (B1), standing up to answer questions (B2) and looking down to read books (B5), and negative behaviors contain two categories of playing with the phone with head down (B3) and looking left and right (B4). At the time of validation, the number of positive and negative behaviors and the percentage of behavioral categories were calculated to provide a data base for the subsequent analysis.

In this paper, we take an interval of 150 seconds to extract the detection results into 20 classroom slices, count the number of students, the number of positive behaviors and the number of negative behaviors in each classroom slice, and compare the results with the manual identification results. Based on the statistical results, the positive behavior image is drawn and the proportion of positive behavior in the whole class is analyzed, and the results are shown in Figure 7. As can be seen from the figure, the recognition results of model recognition and manual recognition in classroom slices are generally consistent, and the model recognition accuracy rate obtained by dividing the number of people accurately recognized in all classroom slices by the total number of people is 94%. Both the results of manual recognition and model recognition show that the proportion of students' positive behaviors in the whole class fluctuates and then decreases as the class progresses. In the first part of the class corresponding to class slices 1-3, the proportion of students' positive behaviors decreased, indicating that students did not fully enter the learning state at the beginning of the class. In the middle of the class, the new knowledge taught by the teacher attracted students' attention and the proportion of positive behaviors increased. As the class progressed to the latter part of the class, the percentage of students' positive behaviors decreased slightly.

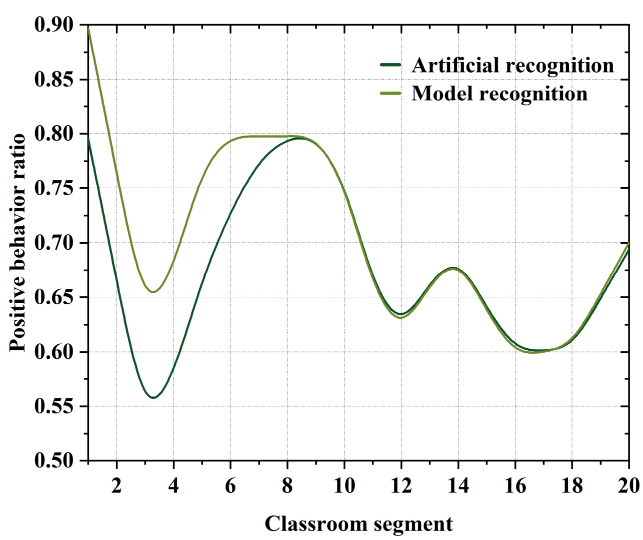


Fig. 7. Comparison of students' positive behavior

In order to further analyze students' behavioral performance in the classroom, this paper feeds the target boxes output from the intermediate process of the student classroom behavior detection model into the student multi-target tracking model, which uses the location information of the target boxes to track the student targets and generates a number for each target to differentiate between students with different identities. This subsection correlates the above number with the classroom behavior categories of students output from the student classroom behavior detection model, and finally obtains the percentage of student behaviors throughout the classroom process, and the results are shown in Fig. 8. The figure shows 10 as the behavioral share of students during the whole English classroom, where the vertical axis is the student number and a set of stacked bars is plotted for each student. Since the behavioral category data of each student comes from each frame of the classroom video, dividing the total number of certain types of behaviors of a student by the total number of all the behaviors of that student can get the percentage of that type of behavior in the whole class, which can also be regarded as the time percentage of that type of behavior. Therefore, in this paper, the horizontal axis of the bar graph is set as a percentage, which can indicate the percentage of each type of behavior of different students in the whole class. From the graph, it can be seen that students numbered 1 to 5 have less than 50% of head-up listening behavior, while students numbered 6 to 10 all have a higher percentage of head-up listening as a positive behavior than the top five students.

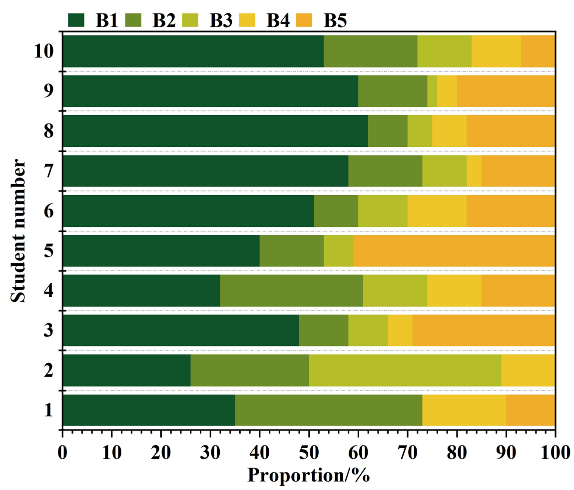


Fig. 8. Student behavior ratio

For further analysis, the study drew a heat map of English classroom behavior of 10 students, and the results are shown in Figure 9. The horizontal coordinate of the heat map represents time, ranging from 0 to 1000 seconds, corresponding to 15 minutes of classroom activity time. The vertical coordinates represent different students, and the students are numbered from 1 to 10. The color change of the heat map represents the degree of positivity of the students' behavior, where red represents positive (1) and blue represents negative (-1). In order to facilitate the observation of the general trend of behavioral changes, the data were smoothed in this paper. The heat map shows that the behavioral pattern of student number 8 shows a higher and more consistent level of positivity throughout the classroom time, which is red for almost the entire period, indicating that the student is more focused and engaged during this time. In contrast, students numbered 1-4 showed more fluctuations in motivation throughout the time period, especially student number 2, whose heat map was blue for most of the time period. Overall, there were several specific moments throughout the class time where all students' motivation decreased, especially at the beginning and end of the class.

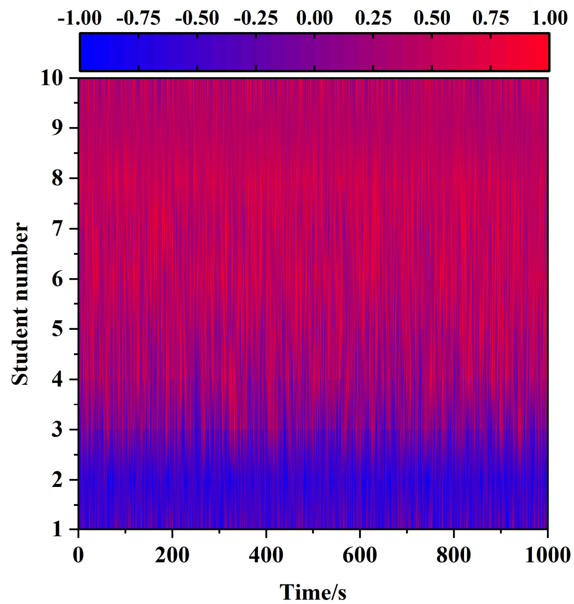
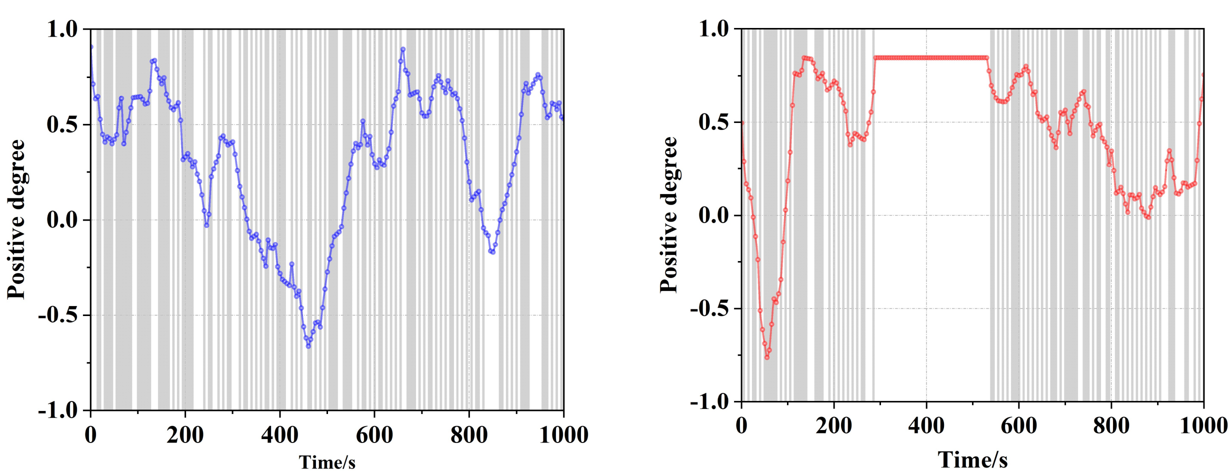


Fig. 9. Students' classroom behavior heat map

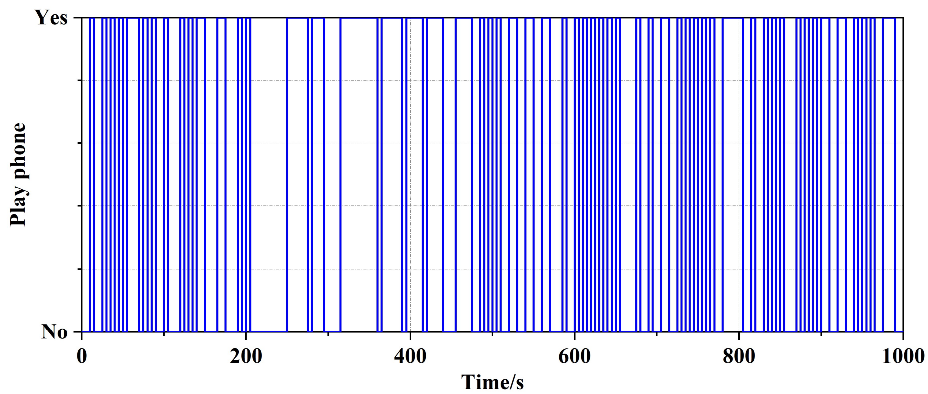
In addition to this, the recognition results of two students numbered 2 and 8 are selected to generate the student state change timing diagrams for further comparative analysis. The comparison graph is shown in Fig. 10, (a) and (b) are the state change time series graphs of student 2 and student 8 respectively. The vertical coordinates in the graph indicate the positive degree of the behavior, with 1 indicating positive and -1 indicating negative. The light gray folds in the graphs are the original data, and the data have been window smoothed and plotted with blue and red folds, respectively, to reflect the trend of temporal changes in behavior. It can be seen that student #2's behavioral positivity shows significant fluctuations, especially in the time period between 270 and 440 seconds, with a continuous downward trend in positivity, which suggests that the student had a low interest in the classroom activity during this period thus leading to inattention. However, this negative trend was not continuous and showed some instability, suggesting that Student #2's attention to class content was intermittent or distracted by matters unrelated to the classroom activity. In contrast, Student #8's behavioral motivation was generally at a high level, and the trend was relatively smooth, with no significant fluctuations. However, the student experienced a short period of lack of motivation at the beginning of the class. During the middle and latter part of the class, between 600 and 900 seconds, although the student's motivation remained at a high level, the change in motivation showed a continuous slow downward trend, and the student's listening state showed some degree of burnout during the middle and latter part of the class.



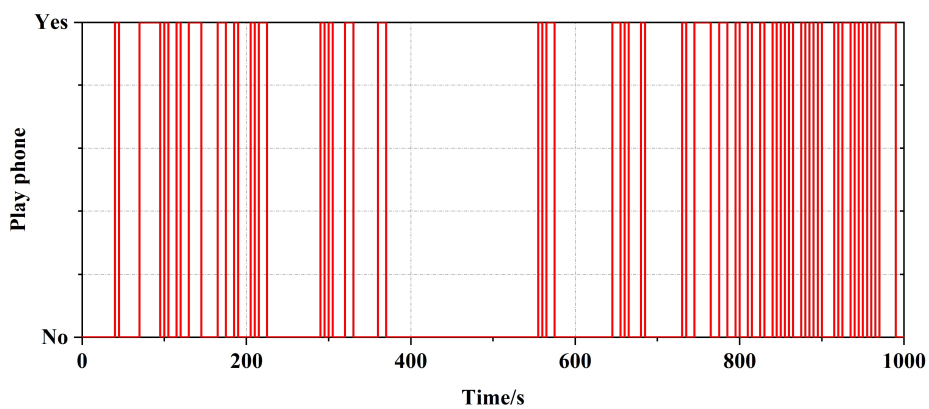
(a) Student 2 (b) Student 8

Fig. 10. Student state change sequence diagram

In the current English classroom in colleges and universities, the behavior of playing cell phones accounts for a high proportion of negative behaviors, so this paper also did a time-series comparative analysis of cell phone playing behaviors of different students, and the results are shown in Fig. 11, with (a) and (b) representing the time-series results of cell phone playing behaviors of Student 2 and Student 8, respectively. The horizontal coordinates of the graph range from 0 to 1000 seconds, representing the duration of the class, and the vertical coordinates are "Yes" for playing cell phone and "No" for not playing cell phone. It can be observed from the graph that Student #2 played with his cell phone several times throughout the class period, and this student may have been distracted by the lack of engaging class content.Student #8's cell phone use was significantly different from that of Student #2.Student #8's cell phone use was more concentrated at specific times in the class, such as the beginning of the class period as well as the middle and end of the class period.



(a) Student 2



(b) Student 8

Fig. 11. Time sequence diagram of students' mobile phone behavior

By visualizing and analyzing the data, teachers can better understand students' learning status and performance, so as to adjust teaching strategies for different classroom phases and improve students' learning participation and learning effects. In addition, education administrators can assess the teaching level of teachers and the learning status of students based on the proportion of students' positive behaviors, as well as analyze the behavioral performance of different classes at the overall level, providing the assistance of AI algorithms to improve the quality of English education in colleges and universities.

5. Conclusion

In this paper, we construct a student behavior analysis method for English classroom based on improved YOLOv7 network and improved DeepSORT multi-target tracking algorithm and other computer vision technologies, which assists English education in colleges and universities to realize intelligent reform and innovative development. The research results show that the recognition rate of the improved YOLOv7 algorithm for all five types of student behaviors reaches more than 90%, and the mAP value of the improved algorithm is improved by 11.12% compared with the original YOLOv7 algorithm. Similarly, the improved DeepSORT multi-target tracking algorithm in this paper improves the MOTA and MOTP metrics by 10.11% and 3.09%, respectively, and the number of ID switches is reduced by 72 times on average, compared with the original DeepSORT algorithm. It is fully demonstrated that the improvement of the two algorithms in this paper is significantly reasonable. Applying this student behavior analysis method to the English classroom and visualizing the results can help teachers grasp the different behaviors of students in different classroom phases, thus assisting teachers in adjusting their teaching strategies according to students' performances and improving the quality of English education in colleges and universities.

References

[1] Lin, Y., Luo, Q., & Qian, Y. (2023). Investigation of Artificial Intelligence algorithms in education. Applied and Computational Engineering, 16, 180-184.

[2] Gligorea, I., Cioca, M., Oancea, R., Gorski, A. T., Gorski, H., & Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. Education Sciences, 13(12), 1216.

[3] Zhao, T., & Cai, Y. (2021). Improvement of English key competences based on machine learning and artificial intelligence technology. Journal of Intelligent & Fuzzy Systems, 40(2), 2069-2081.

[4] Qian, L. (2022). Research on college english teaching and quality evaluation based on data mining technology. Journal of Applied Science and Engineering, 26(4), 547-556.

[5] Hockly, N. (2023). Artificial intelligence in English language teaching: The good, the bad and the ugly. Relc Journal, 54(2), 445-451.

[6] Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. Educational Technology & Society, 26(1), 112-131.

[7] Li, X. (2022, August). Research on the application of computer artificial intelligence technology in feedback teaching of English. In 2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA) (pp. 1023-1027). IEEE.

[8] Schmidt, T., & Strasser, T. (2022). Artificial intelligence in foreign language learning and teaching: a CALL for intelligent practice. Anglistik: International Journal of English Studies, 33(1), 165-184.

[9] Qi, S., Liu, L., Kumar, B. S., & Prathik, A. (2022). An English teaching quality evaluation model based on Gaussian process machine learning. Expert Systems, 39(6), e12861.

[10] Zhang, F., & She, M. (2021). Design of English reading and learning management system in college education based on artificial intelligence. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-10.

[11] Al-Kadi, A., & Ali, J. K. M. (2024). A Holistic Approach to ChatGPT, Gemini, and Copilot in English Learning and Teaching. Language Teaching Research Quarterly, 43, 155-166.

[12] Xin, D., & Shi, C. (2024). Application of Hybrid Image Processing Based on Artificial Intelligence in Interactive English Teaching. ACM Transactions on Asian and Low-Resource Language Information Processing.

[13] Rangel-de Lazaro, G., & Duart, J. M. (2023). You can handle, you can teach it: Systematic review on the use of extended reality and artificial intelligence technologies for online higher education. Sustainability, 15(4), 3507.

[14] Xu, J. (2023). Big data-driven English teaching for social media: a neural network-based approach. Evolutionary Intelligence, 16(5), 1589-1597.

[15] Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. British Journal of Educational Technology, 55(6), 2503-2529.

[16] Özdere, M. (2023). The integration of artificial intelligence in English education: Opportunities and Challenges. Language Education and Technology, 3(2).

[17] Dai, K., & Liu, Q. (2024). Leveraging artificial intelligence (AI) in English as a foreign language (EFL) classes: Challenges and opportunities in the spotlight. Computers in Human Behavior, 108354.

[18] Yang, C., Huan, S., & Yang, Y. (2020). A practical teaching mode for colleges supported by artificial intelligence. International Journal of Emerging Technologies in Learning (IJET), 15(17), 195-206.

[19] Li, Y., & Han, L. (2024). The role of big data and artificial intelligence in the reform and innovation of intelligent English teaching. Journal of Computational Methods in Sciences and Engineering, 24(6), 3341-3353.

[20] Ghafar, Z. N., Salh, H. F., Abdulrahim, M. A., Farxha, S. S., Arf, S. F., & Rahim, R. I. (2023). The role of artificial intelligence technology on English language learning: A literature review. Canadian Journal of Language and Literature Studies, 3(2), 17-31.

[21] Fitria, T. N. (2021). The use technology based on artificial intelligence in English teaching and learning. ELT Echo: The Journal of English Language Teaching in Foreign Language Context, 6(2), 213-223.

[22] Chai, Y., & Liu, N. (2022). Application of artificial intelligence technology in college english teaching under the background of big data. In 2021 International Conference on Big Data Analytics for Cyber-Physical System in Smart City: Volume 1 (pp. 543-555). Springer Singapore.

[23] Mukhallafi, T. A. (2020). Using artificial intelligence for developing English language teaching/learning: an analytical study from university students' perspective. International Journal of English Linguistics, 10(6), 40.

[24] Tiwari, H. P. (2023). An Optimization-Based Artificial Intelligence Framework for Teaching English at College Level Under Tribhuvan University. Journal of English Language Teaching and Linguistics, 8(1), 79-91.

[25] Wang, Y. (2023). Artificial intelligence technologies in college English translation teaching. Journal of psycholinguistic research, 52(5), 1525-1544.

[26] Dong, Y. (2022). Application of artificial intelligence software based on semantic web technology in english learning and teaching. Journal of Internet Technology, 23(1), 143-152.

[27] Tong, Q., & Liu, S. (2025). Intelligent Classroom Environment: Application of Internet of Things and Artificial Intelligence in English Teaching. International Journal of High Speed Electronics and Systems, 2540364.

[28] Simona Tiribelli,Benedetta Giovanola,Rocco Pietrini,Emanuele Frontoni & Marina Paolanti. (2024). Embedding AI ethics into the design and use of computer vision technology for consumer's behaviour understanding. Computer Vision and Image Understanding,248,104142-104142.

[29] Qingqi Wu,Lihui Cen,Shichao Kan,Yongping Zhai,Xiaofang Chen & Hong Zhang. (2025). Real-time underwater target detection based on improved YOLOv7. Journal of Real-Time Image Processing, 22(1),43-43.

[30] Kai Zhu,Junhao Dai & Zhenchao Gu. (2024). Dynamic Tracking Method Based on Improved DeepSORT for Electric Vehicle. World Electric Vehicle Journal,15(8),374-374.