Optimization of Student Behavior Detection Algorithm Based on Improved SSD Algorithm

Yongqing CAO, Dan LIU
College of Computer Science and Engineering
Cangzhou Normal University
Cangzhou, 061001, Hebei, China

***Abstract*—Despite advancements in educational technology, traditional action recognition algorithms have struggled to effectively monitor student behavior in dynamic classroom settings. To address this gap, the Single Shot Detector (SSD) algorithm was optimized for educational environments. This study aimed to assess whether integrating the Mobilenet architecture with the SSD algorithm could improve the accuracy and speed of detecting student behavior in classrooms, and how these enhancements would impact the practical implementation of behavior-monitoring technologies in education. An improved SSD algorithm was developed using Mobilenet, known for its efficient data processing capabilities. A dataset of 2,500 images depicting various student behaviors was collected and enhanced through preprocessing methods to train the model. The optimized SSD model outperformed traditional algorithms in accuracy and speed, thanks to the integration of Mobilenet. Evaluation metrics such as precision, recall, and frames per second (fps) confirmed the superior performance of the Mobilenet-enhanced SSD algorithm in real-time environmental analysis. This advancement represents a significant improvement in surveillance technologies for educational settings, enabling more precise and timely assessments of student behavior. Despite the promising outcomes, the study faced limitations due to the uniformity of the dataset, which mainly consisted of controlled environment images. To improve the generalizability of the findings, it is suggested that future research should broaden the dataset to encompass a wider range of educational settings and student demographics. Additionally, it is encouraged to explore alternative advanced machine learning frameworks and conduct longitudinal studies to evaluate the influence of real-time behavior monitoring on educational outcomes.**

***Keywords*—Improved single shot detector (SSD) model; mobilenet network; class behavior recognition; artificial intelligence**

NOMENCLATURE TABLE

Identifier	Description
Abbreviations	
SSD	Single Shot Detector
AI	Artificial Intelligence
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
Mobilenet	Mobile Networks
VGG16	Visual Geometry Group Network 16
DIoU-NMS	Distance Intersection over Union - Non-Max Suppression
Symbols	
x	Input data
y	Output data
σ	Activation function
α	Learning rate
β	Regularization parameter
θ	Parameters of the model
W	Weight matrix
b	Bias terms
Greek Symbols	
γ	Discount factor in reinforcement learning
δ	Difference or error term in calculations
λ	Weight decay factor
Subscripts	
i	Index for summation
j	Index in a series or layer
t	Time step index in sequences
Superscripts	
$-$	Denotes the previous state in recursive formulas
+	Denotes next state in progressive calculations

I. INTRODUCTION

The current higher vocational education reform is increasingly emphasizing hybrid teaching as the focal point and direction of development, owing to the rapid advancement of information technology and the deepening of teaching information reform. By integrating network information technology into classroom instruction, it enhanced teaching quality. This work done by Huang et al. [1] focused on reforming the Modern Educational Information and Technology course through both online and offline methods. Through analysis of 50 questionnaires, it shed light on the efficacy of this approach. In the realm of education, real-time insights into student learning were offered by surveillance videos, yet limitations were faced by current action recognition methods. To address this, a novel dataset was created from smart classrooms, notable for its complex backgrounds and crowded scenes. The solution suggested by Li et al. included an attention-based relational reasoning module and a relational feature fusion module, enhancing recognition accuracy. Through rigorous experimentation, existing algorithms were surpassed by our model, signaling a new era of action recognition in education [2]. Trabelsi et al. [3] advocated for AI-powered classrooms that monitored student attention, even with face masks. Their study refined the YOLOv5 model, achieving a 76% average accuracy. They argued this technology empowered instructors to craft tailored learning experiences, blending tradition with innovation in education. The paper done by Tran et al. [4] discussed the application of computer vision in education to monitor and analyze student behavior. It proposed a new method that used the movement of students' body parts to identify classroom behaviors, with a database of ten actions for method evaluation. A deep learning model enabled real-time action analysis and classification, showing effective results in enhancing teaching by providing feedback for lesson adjustment based on student engagement.

In their scholarly endeavor, Park and Kwon [5] crafted a potent educational program that seamlessly integrated artificial intelligence (AI) into South Korea's middle school free semester system. Through meticulous preparation, development, and improvement, they honed a curriculum focused on technology education's specific needs. Their program, distinguished by its emphasis on AI's societal impact, ethical considerations, and problem-solving, yielded remarkable outcomes. Students exhibited increased interest in technology, aspirations for technological careers, and enhanced understanding of AI's implications. Park and Kwon's study served as a beacon, illuminating the path for future educators to infuse AI into technology education effectively. In their seminal work, Sharma et al. [6] highlighted the transformative potential of artificial intelligence (AI) and machine learning (ML) in education. They advocated for AI's role in creating personalized learning content, analyzing student data to enhance teaching strategies, and automating grading and feedback processes. Through these innovations, education became more effective, personalized, and engaging, promising a brighter future for learners and educators alike.

The detection of students' classroom degrees is based on the target recognition algorithm, also composed of artificial intelligence. Testing the students' classroom behavior through the target algorithm can efficiently determine the total count of students present within the classroom and have more accurate statistical results. Face recognition in the class can also be done in a short time. Combined with the above content, it is highly feasible to analyze students' classroom behavior through artificial intelligence.

The crucial enhancement of the quality of education through the integration of information technology in classroom instruction is being addressed as technology continues to advance rapidly. Challenges are faced by traditional action recognition algorithms in the complex and crowded environments of smart classrooms. An optimized Single Shot Detector (SSD) algorithm is presented in this study, specifically designed to improve the accuracy of detecting student behavior. Through the incorporation of an attention-based relational reasoning and feature fusion module, the refined model not only overcomes the limitations of existing methods but also enhances the real-time analysis of student engagement. The objective of our research is to validate this approach by demonstrating its superior performance in detection accuracy through extensive testing against traditional models. Valuable insights for the integration of AI-driven tools in education will be provided by this validation.

The main Contributions of the present work are as follows:

1) Improved SSD algorithm: Integrated with MobileNet, specifically designed for dynamic educational environments to enhance real-time precision and speed in identifying student actions.

2) Tailored dataset creation: Consists of 2500 images showcasing a variety of student behaviors, customized for training in authentic classroom scenarios.

3) Innovative data processing methods: Incorporates feature fusion and attention-driven relational reasoning, enhancing the accuracy and efficiency of behavior analysis.

4) Instantaneous environmental evaluation: The incorporation of MobileNet elevates the SSD model's capacity for prompt and efficient classroom surveillance.

This paper is organized as follows: The paper begins with an introduction in Section I and is followed by giving related work in Section II, highlighting advancements and identifying gaps in classroom state identification and target detection algorithms, emphasizing the integration of deep learning in educational settings. In the Student Classroom Behavior Recognition section, the development of the enhanced Single Shot Detector (SSD) algorithm and the creation of a novel dataset are presented, focusing on data collection and image enhancement. Also, technical modifications, including low-level feature enhancements and the transition to a more efficient Mobilenet model are provided in Section III. Section IV outlines testing conditions and evaluation metrics, compares the model against traditional algorithms, and Section V presents the results and discussion on the effectiveness of our optimized algorithm in real-time student behavior detection and its implications for educational technology. The paper concludes in Section VI with a Conclusion, summarizing contributions and exploring future research directions.

II. Related Work

A. A Review of the Literature on Classroom State Identification

The study of students' classroom state abroad began earlier than in China. One of the earliest articles on the status of the classroom for students was published in 1962. Through searching statistics on the Internet, it was found that foreign research on students' classroom status was concentrated from 1998 to 2017, among which the research on student concentration reached a climax from 2006 to 2024.

China lags behind foreign countries in the research of students' classroom category. Based on the statistical evaluation of CNKI, the classroom status research of students only covers about a quarter of the foreign countries; it is also relatively late, mainly between 2012 and 2024. Much of this literature is about the research on cultivating students' good classroom state rather than studying the "classroom state." With the development of deep learning in recent years, using deep learning has gradually appeared to conduct related research on students' classroom behavior, state, fatigue degree, and other aspects. For example, Hasnine et al. developed a classroom monitoring system within the MOEMO framework for online courses, visualizing students' emotional states. This allowed teachers to intervene promptly when students were disengaged, improving overall engagement and concentration, and optimizing instructional strategies [7].

B. A Literature Review on Target Detection Algorithms

In their paper, Shuai and Wu [8] introduced enhancements to the SSD object detection algorithm, integrating Batch Norm operation for improved generalization and faster training. They also incorporated object counting functionality into image recognition within the SSD framework. Their implementation of a detection system, utilizing Flask and Layui frameworks, enabled real-time selection and display of detection results on the front-end interface. Hu et al. [9] presented a novel approach to sea urchin detection, tackling the shortcomings of the classic SSD algorithm. Their feature-enhanced method combined multidirectional edge detection and integration of ResNet 50, achieving an impressive 81.0% Average Precision (AP) value—an improvement of 7.6% over SSD. Tested on the National Natural Science Foundation of China's underwater dataset, their algorithm proved effective in accurately detecting sea urchins, particularly small targets, heralding a new era in autonomous aquatic exploration. Yan's [10] research focused on the ever-evolving field of computer vision-based motion target detection and tracking. Through the enhancement of traditional approaches and the introduction of novel fusion techniques, Yan was able to increase detection accuracy by 2.6% without compromising real-time efficiency. By carefully adjusting parameters, the system attained both stability and precision, marking a significant advancement in the realm of surveillance and interaction technologies.

Liu's research [11] introduced a transformative method for evaluating the learning progress of English students in higher vocational colleges. By enhancing the Single Shot MultiBox Detector (SSD) algorithm, Liu expanded the capabilities of detection and improved its accuracy. The approach utilized Multi-Task Convolutional Neural Networks (MTCNN) and multi-level reduction correction, resulting in promising outcomes. The mean deviation was below 1.5, and the accuracy exceeded 90% across various student behaviors. Liu's work exemplified the fusion of academic inquiry and technological innovation, providing a practical solution for precise and reliable assessment of student's status in educational environments. In a similar study, Zhang and Xu [12] creatively combined deep learning algorithms with teacher monitoring data to develop the MobileNet-SSD, refining English classroom instruction. Despite initial challenges, their optimization efforts yielded remarkable results. The algorithm achieved an average detection accuracy of 82.13% and a rapid processing speed of 23.5 frames per second (fps) through rigorous experimentation. Notably, it excelled in identifying students' writing behaviors with an accuracy rate of 81.11%. This advancement not only enhanced the recognition of small targets without compromising speed but also surpassed previous algorithms, promising modern technical support for English teachers and improving the efficiency of classroom teaching. Wang et al. [13] introduced C-SSD; an enhanced small target detection method based on improved SSD architecture. By replacing VGG-16 with C-DenseNet and incorporating residuals and DIoU-NMS, C-SSD achieved superior accuracy, outperforming other networks with an impressive 83.8% accuracy on the PASCAL VOC2007 test set. Notably, C-SSD struck a fine balance between speed and precision, showcasing exceptional performance in swiftly detecting small targets, marking a significant advancement in target detection technology. Nandhini and Thinakaran [14] proposed a novel approach to object detection, addressing the challenges of identifying small, dense objects with geometric distortions. Their deformable convolutional network with adjustable depths blended deep convolutional networks with flexible structures, yielding superior accuracy in recognizing objects. Experimental validation confirmed significant improvements in accuracy, highlighting the framework's potential to enhance machine vision capabilities in complex visual environments.

Cheng et al. [15] addressed the challenge of accurately and rapidly detecting concealed objects in terahertz images for security purposes. Their novel method enhanced the SSD algorithm with a deep residual network backbone, a feature fusion-based detection algorithm, a hybrid attention mechanism, and the Focal Loss function. Results showed a significant accuracy improvement to 99.92%, surpassing mainstream models like Faster RCNN, YOLO, and RetinaNet, while maintaining high speed. Their approach offered valuable insights for the application of deep learning in terahertz smart security systems, promising real-time security inspections in public scenarios. Dai [16] proposed an online English teaching quality evaluation model that combined K-means and an improved SSD algorithm. DenseNet replaced the backbone network for enhanced accuracy, while quadratic regression addressed sample imbalance. A feature graph scaling method and k-means clustering optimized default box parameters. Utilizing a dual-mode recognition model, Dai predicted students' states during teaching, demonstrating superior detection accuracy compared to alternative algorithms.

Despite the prevalence of research on target detection using deep learning both domestically and internationally, there is a scarcity of studies specifically addressing the detection of students' classroom behavior using these technologies. This paper aims to bridge this gap by optimizing the Single Shot Detector (SSD) algorithm through a comprehensive review of relevant literature and adapting it to real classroom environments. The main objective of this research is to develop and improve a dataset for classroom behavior. Given the limited availability of images, we employ random enhancement techniques such as translation, noise addition, and color adjustment to fulfill the training requirements. This dataset encompasses scenarios where students in the back rows are detected as small objects using low-level features of the SSD algorithm. To enhance object recognition accuracy, we integrate both shallow and deep information layers.

Furthermore, we address the limitations of the traditional SSD algorithm, which relies on the VGG16 network known for its extensive parameters that impede processing speed and demand high computational power. By transitioning to an enhanced Mobilenet model that incorporates network depth and separable convolutions, we significantly reduce the parameter load while maintaining robust classification capabilities, thereby improving recognition speed.

The integration of technology in education has presented a notable obstacle in accurately evaluating classroom dynamics, particularly in complex and crowded environments. In order to address this challenge, this research study introduces a tailored Single Shot Detector (SSD) algorithm that is specifically designed for educational settings. Conventional action recognition algorithms often encounter difficulties in accurately monitoring student behavior in dynamic classroom settings. To overcome these limitations, our study incorporates advanced target recognition algorithms and utilizes a distinct dataset obtained from smart classrooms. This optimized SSD algorithm aims to offer real-time and accurate analyses of student engagement, signifying a significant advancement in the implementation of educational technology.

III. Student Classroom Behavior Recognition Algorithm Based on an Improved SSD Algorithm

A. *The Construction Process of the Classroom Behavior Recognition Model*

Classroom behavior helps analyze the quality of students' lectures and the teaching effect. Therefore, this paper chooses five common classroom gestures, namely, sitting and listening, raising hands, writing, sleeping, and playing with a cell phone, to identify and study. This chapter analyzes the shortcomings of SSD by target detection algorithm, proposes the improved SSD algorithm combined with the characteristics of students' classroom behavior, and provides a detailed introduction to the behavior recognition model's application procedure in the class. For target detection, pre-processing of data, training data, and other processes are required. The analysis process is explained in Fig. 1 below in more detail [17]. Peng's seminal work [18] emphasized the importance of precise pronunciation in English teaching. Introducing a novel clustering-enhanced SSD algorithm, the paper addressed limitations in pronunciation detection, enhancing feature extraction and detection speed. By integrating multiscale features and channel attention mechanisms, it improved accuracy while reducing computation. Employing K-means clustering optimized parameter settings, yielding precise evaluation of oral English proficiency. This pioneering approach marked a significant stride in the fusion of technology and pedagogy, promising a future of unparalleled linguistic mastery.

In this topic, detecting students 'classroom behavior requires building a data set first. In this data, 2,500 pictures of students' classroom behavior were obtained through the network and shooting methods, including the students' behaviors of raising their hands and standing up in class and the state of sleeping and writing. In the above five students' classroom behaviors, the number of pictures of each behavior was 500. Subsequently, the collected data is collated by building the database, and the test, training, and validation set is formed. Finally, the data in the three sets are measured. That is, the data in the data set is input into the model, and the research results are obtained by comparing them with the verification set and whether the expectations can be determined through the analysis of the results. If the accuracy and feasibility of the model after this experiment are relatively high, the model is still retained, and the validation of other similar studies is completed [19].

B. *Principles of the SSD Algorithm*

The process of detecting students' classroom behavior is optimized based on SSD detection. The basic detection algorithm is first described below. Different SSDs are divided into SSD300 and SSD512 according to the input image size. After entering the data set, the data with the image size of SSD300 is extracted. This type of network structure is realized through the basic network, in which the image is subsequently processed through the neural network to complete the feature extraction and selection of the data. After processing the VGG16, the proposed part can be supplemented by adding the convolution level. The specific process is shown in Fig. 2.

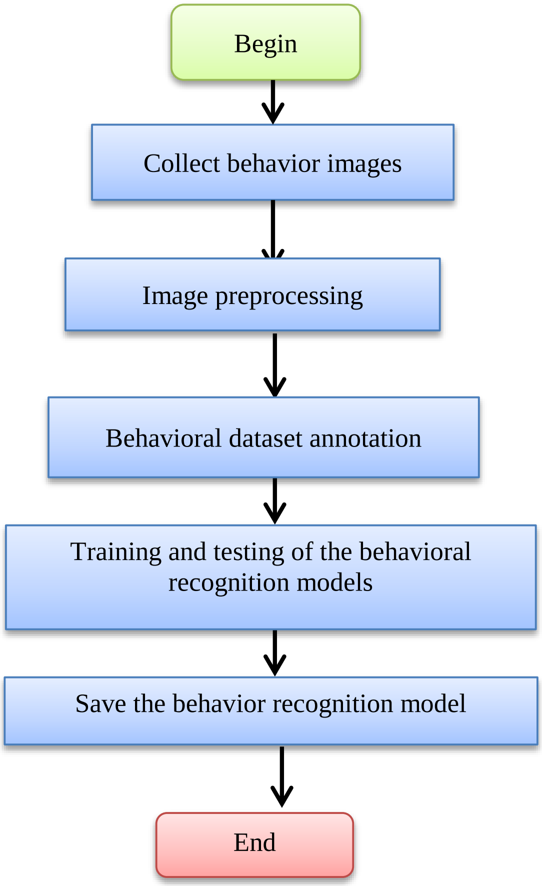

Begin
Collect behavior images
Image preprocessing
Behavioral dataset annotation
Training and testing of the behavioral recognition models
Save the behavior recognition model
End

Fig. 1. Classroom behavior recognition process.

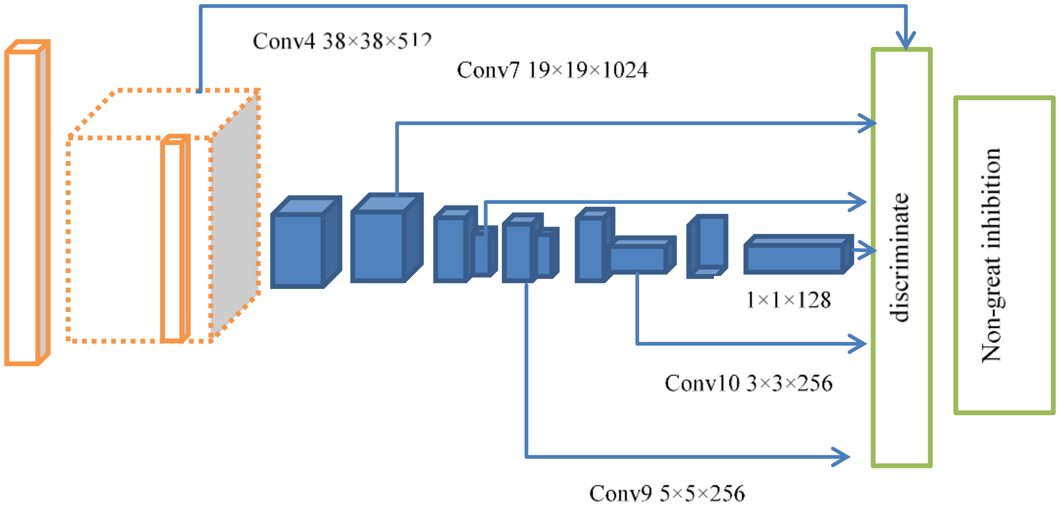


Fig. 2. The SSD network structure.

According to the previous content, SSD and POLO are target detection algorithms of one-stage type; however, their feature extraction methods are different. The early POLO algorithm only extracted the information of the highest-level features through the convolution operation, so although the semantics are high, the small target information may need to be recovered. Therefore, as previously stated, the early POLO algorithm is fast, but the small target detection rate is not high. The SSD algorithm uses a variety of scale feature graph detection. After the modified VGG16 basic network increases gradually, decreasing the convolution layer, and then selected six layers from all layers, their size from before to back is reduced, where the feature graph size is used to identify small objects, features of small graph size is used to identify large objects. In doing so, image features can be obtained from different levels to get shallow information and extract deeper, more abstract information.

C. *Improvement of the SSD Algorithm*

The SSD algorithm's structure is described above, indicating that the algorithm is mainly based on the feature extraction of the image and then obtains the detection results by detecting the feature layer. Although some scholars, through the algorithm to detect the results of the target, found that its feasibility is strong, on the whole, the SSD algorithm is still based on the basic network as the premise for better classification of data processing, but due to the uncertainty of data quantity, for the data amount of relatively large parameter processing performance is insufficient. For example, after removing the full connection layer of the algorithm, the resulting parameter is 14122995, and about 3 / 2 of the time, it is conducted in the basic network. Thus, training is more challenging through the proposed algorithm. On the other hand, due to the structural influence of the traditional SSD algorithm, some data are carried out through the shallow layer. Still, the shallow layer of information is relatively insufficient, and it is difficult to achieve the expected detection effect. Combined with the above discussion, the traditional SSD algorithm is optimized to improve the accuracy and feasibility of the research results. After various algorithms, the lightweight network is adopted instead of VGG16 to reduce the effect of the number of parameters on the training results and improve the accuracy and efficiency of target detection. The process of optimization is described below.

1) SSD infrastructure network improvements: According to the optimization process above, it can be seen that the detection of students' classroom behavior in this topic is replaced by eliminating VGG16 based on excluding VGG16 with fewer parameters. After the analysis of relevant data, the network meets the standard. The optimized algorithm saw the original number of 13.3 million parameters for up to 4.2 million, greatly saving the training time. After training the network dataset, it is found that the optimized algorithm can greatly improve the detection efficiency, and the optimized algorithm is slightly lower and negligible. Considering the above content, the research on students' classroom behavior in this topic is formally detected based on the optimization algorithm. An overview of the network optimization process is given in Table I.

TABLE I. COMPARISON TABLE

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
Mobilenet (244)	70.6%	569	4.2
VGG16	71.5%	15300	138

After changing the parameter value, it is found that the reason for the budget efficiency improvement and the decrease in the parameter value is that the CNN in the network composition is in a separable state, and the main part is the deeply separable convolution. If the hierarchy describes the optimized algorithm, the network results on both sides can be expressed as 28 layers, while the network results on one layer are expressed as 14. The CNN operation process is explained above and will not be repeated here. After putting the image of students' classroom behavior into input, the feature extraction based on the CNN can obtain a feature information map and then process the data through BN and ReLu, train other images with a CNN, and obtain the operation results of the above two operations. The depth of convolution and point convolution of the operation process is shown in Fig. 3.

This paper improves the Mobilenet in two aspects: In the network structure diagram given in Fig. 3, it is evident that after every point convolution or deep convolution is completed, the search needs to be followed by an activation function and a normalization method using ReLu and BN, respectively. However, in the article on the BN layer, fully connected layer, and convolutional layer relationship, these three are linear relations, so combining the BN layer into either will have little impact on the results. The efficiency of the BN layer calculation was enhanced by integrating it with the preceding convolution, thereby resulting in enhanced speed based on the previous foundation. Adjusting the Mobilenet input size involves modifying it from 224X224 to 300X300. This modification serves two purposes: initially, enlarging the input size has the potential to augment the capacity of feature map information, consequently improving detection accuracy. However, it is important to note that increasing the input size excessively significantly escalates network parameters, compromising the model's lightweight nature. In the present study, the SSD network structure utilizes an input size of 300X300, laying the groundwork for the subsequent amalgamation of the two networks.

According to the basic network results, the dynamic data is intercepted, and the image features are processed by the optimization algorithm mentioned above, along with the image features, to obtain the feature image information. This topic trains the students' classroom behavior detection by training the ordinary-size convolutional layer connection and then understanding the deep image information through its results. The image information obtained through the algorithm is placed in the classification for judgment, and the data is not regressed according to the traditional algorithm. The completion of the replacement of the fundamental network has been achieved. In the end, a selection of six feature layers has been made, just like the original SSD, to accomplish the task of feature extraction and target detection. The selection process should take into account the depth of the layer. In the event that the depth is insufficient, it becomes challenging to extract an adequate amount of image information. Thus, the six feature layers selected in this paper decrease anterior to posterior size for multiscale prediction. In this step, the improved part of the Mobilenet network is combined with the SSD network framework to obtain the new network. Fig. 4 visually represents the updated network structure.

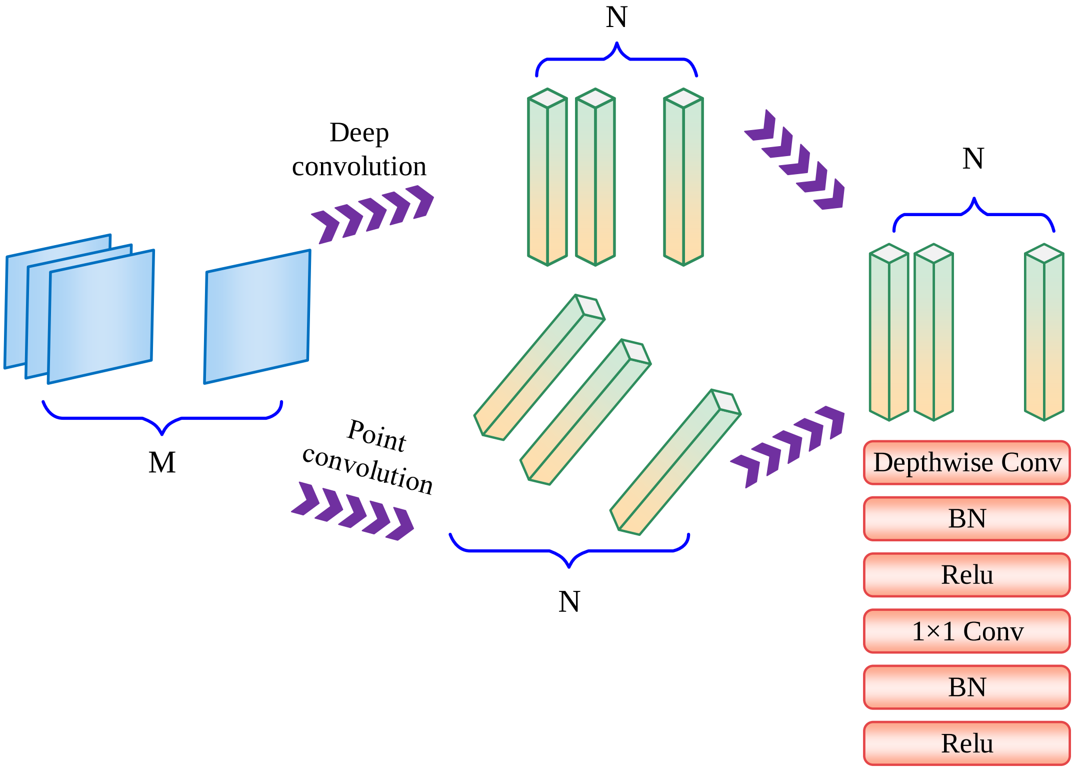


Fig. 3. Depth-separable convolution model.

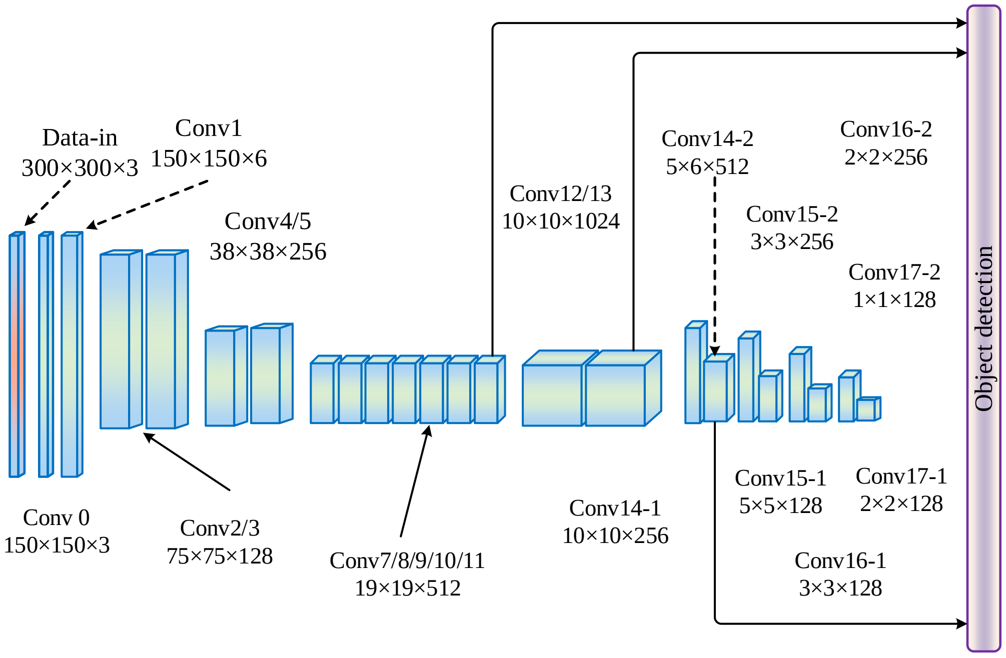


Fig. 4. Enhanced network structure.

2) *Feature fusion of the network models*: In the final stage, replacing the fundamental network enhanced the detection speed, albeit without any noticeable improvement in the accuracy of detecting small targets. A widely employed strategy for enhancing the model's performance involves the integration of features at various scales. Thus, this section introduces the approach of feature fusion and proposes the model fusion strategy accordingly. Based on the characteristics of the network model structure and feature fusion method obtained in Section III(C)(1), the feature fusion method chosen is the additive approach to integrate the network. Among the six characteristic layers extracted from the model structure, the dimensions progressively decrease from shallow to deep, with less abstract information being presented initially. Transferring the abstract data from the deep feature layer to the shallow layer is the goal of feature fusion. The structure after the network fusion is shown in Fig. 5, using the fusion feature layer for feature extraction and detection operations. After feature fusion, low-level feature maps can contain high-level information to enhance the detection effect of small targets and improve detection accuracy. The dimensionality of the network remains unchanged after the operation described above and remains six feature layers for detection.

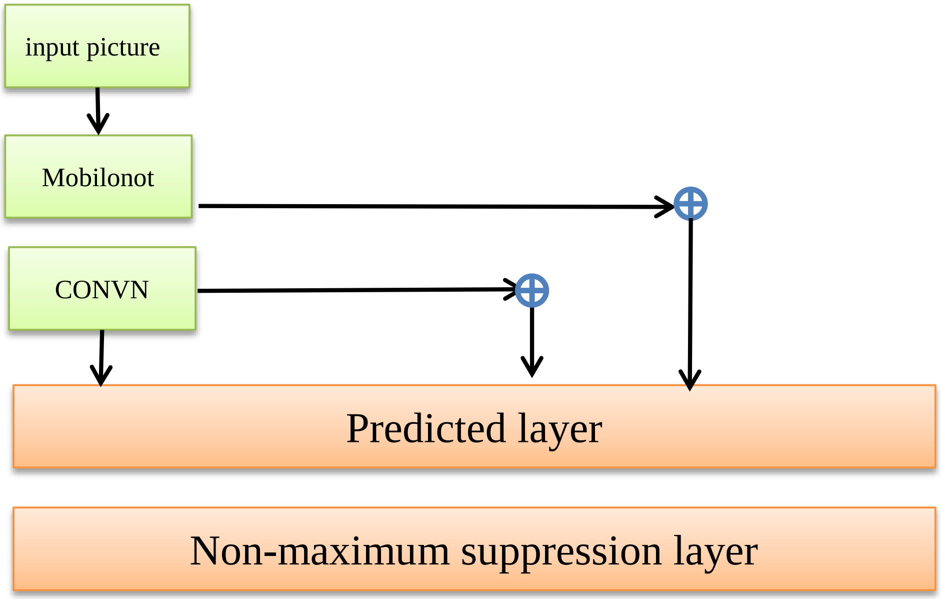


Fig. 5. The improved network structures.

3) The RMSProp optimization algorithm: The algorithm measures the historical gradient across all dimensions to find the square and then superposition while introducing the decay rate, yielding a historical gradient sum. The detection results and learning rate of the image features are calculated, and the detection results' accuracy and efficiency are improved through the optimization algorithm proposed in this paper. The calculation formula is as follows: Eq. (1) and Eq. (2):

$$S_{dR} = \beta S_{dR} + (1-\beta)(dR)^2 \quad (1)$$

$$R = R - \rho \frac{dR}{\sqrt{S_{dR}+a}} \quad (2)$$

IV. Experiment and Testing

This study primarily examines the conventional SSD algorithm, the unenhanced Mobilenet-SSD algorithm, and the enhanced Mobilenet-SSD algorithm by considering three key factors: training complexity, detection accuracy, and detection velocity. The difficulty of training refers to the value of the three model loss functions in the same training time and training times.

The following quantitative analysis compares the three algorithms to verify the detection accuracy of SSD, Mobilenet-SSD, and the proposed algorithm. The three algorithms above are assessed using the same experimental conditions. The data of the experimental environment are shown in Fig. 6. To facilitate the efficiency of the experiment, the verification process is processed by a self-made data set, and the accuracy and efficiency of students' second classroom behavior detection obtained by comparing the three algorithms are used as the evaluation standard.

According to the data analysis results in Fig. 6, in the identification of students 'classroom behavior in the three algorithms, the accuracy of the Mobilenet-SSD algorithm and SSD algorithm is 76.14% and 84%, respectively, and the accuracy of the optimization algorithm proposed in this paper is 85.21%. It can be seen that the optimized algorithm can more accurately identify students' classroom behavior. On the other hand, in terms of detection efficiency, the time of the first two algorithms is 27.1 and 22, respectively. In contrast, the detection speed of the optimization algorithm is relatively high, and its value is 21. Combined with the above two detection contents, the optimized algorithm is more accurate for students' classroom behavior detection, and the detection speed is faster. The difference in the loss function's change curve during the training process can be used to determine how difficult the model is to train. The Mobilenet-SSD and SSD models were performed on the loss function curves with 100 iterations of 50,000 times, as shown in Fig. 7.

Evidently, both models' loss values are visible and are always declining, indicating that both models are more reasonable. In the training time, the loss dropped to 0.5; the model used in this paper took about six days, and the original SSD model took about eight days. Furthermore, the figure illustrates that the rate at which the loss value decreases for this model is higher, indicating that the training process for the model employed in this study is comparatively easier than that of the conventional SSD model. The optimization algorithm proposed in this topic is optimized on the basis of the traditional SSD model to detect students' five common class behaviors. The results of the algorithm detection are shown in Fig. 8.

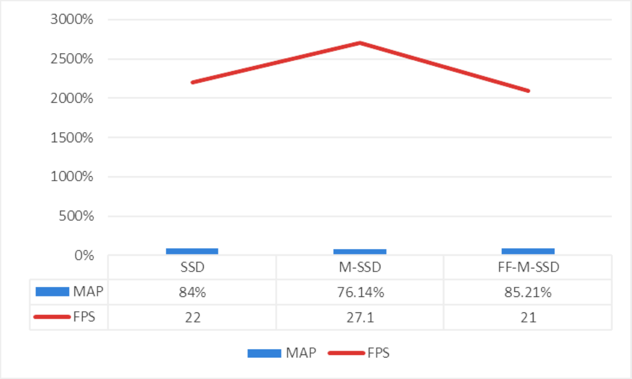


Fig. 6. Different model identification effect.

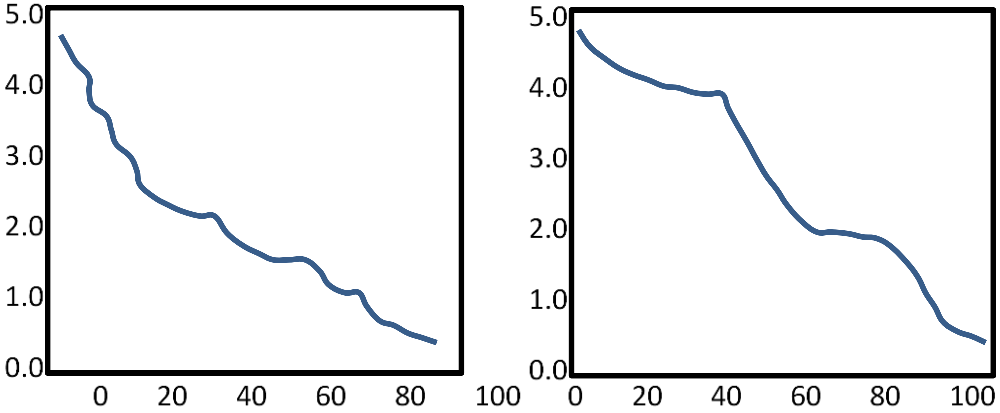
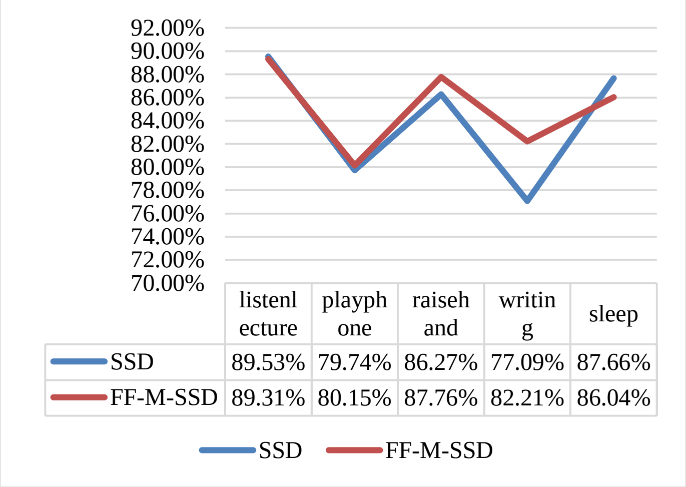


Fig. 7. Loss.



	listenlecture	playphone	raisehand	writing	sleep
SSD	89.53%	79.74%	86.27%	77.09%	87.66%
FF-M-SSD	89.31%	80.15%	87.76%	82.21%	86.04%

Fig. 8. Comparison of students' classroom behavior detection algorithm.

Based on the initial SSD algorithm, the Mobilenet-SSD algorithm has enhanced the detection performance of small objects across all five actions. Notably, the improvement in recognizing writing has reached 3.03%, underscoring the model's advancement in small object recognition. Observing the Mobilenet-SSD identification results of feature fusion, it was found that the movement detection accuracy was the highest, and the movement of writing and playing with the mobile phone was the lowest. After collecting relevant data and analyzing it, it is believed that the main reason for the above situation is that in the student's classroom behavior, the action is easily confused during the identification period, which leads to a relatively weak detection effect.

Table II shows the comparison results for different metrics and various versions of SSD.

TABLE II. COMPARISON RESULTS FOR DIFFERENT METRICS AND VARIOUS VERSIONS OF SSD

Metric	Conventional SSD	Unenhanced Mobilenet-SSD	Enhanced Mobilenet-SSD
Training Complexity (Time to reach loss=0.5)	~8 days	N/A	~6 days
Detection Accuracy	84%	76.14%	85.21%
Detection Velocity (Units)	27.1 units	22 units	21 units

The findings of this study validate the significant advancements made by the improved Mobilenet-SSD algorithm in terms of detection accuracy and speed. Additionally, it demonstrates a noteworthy decrease in training complexity when compared to traditional SSD and unenhanced Mobilenet-SSD algorithms.

Table III highlights the distinctions and contributions of the present study compared to existing related work, with a focus on educational technology and AI.

TABLE III. DISTINCTIONS AND CONTRIBUTIONS OF THE PRESENT STUDY COMPARED TO EXISTING RELATED WORK

Comparison Criteria	Present study	Existing Related Work
Research Focus	Optimization of the SSD algorithm integrated with MobileNet specifically for real-time student behavior detection in educational settings.	General improvements in action recognition algorithms for varied applications, including but not limited to educational settings.
Dataset Customization	Development of a novel dataset from smart classrooms, designed to capture complex student behaviors specific to educational environments.	Use of broader, less specific datasets primarily focused on general object or action recognition that may not address the unique challenges of educational settings.
Performance Optimization	High emphasis on both detection accuracy (85.21%) and processing speed, suitable for real-time educational applications.	Studies often emphasize either accuracy or speed but may not balance both, particularly not in the context of real-time educational needs.
Technological Innovations	Implementation of feature fusion techniques and lightweight deep learning architectures tailored to the specific needs of monitoring classroom dynamics.	Application of existing deep learning models (e.g., YOLOv5, classic SSD) often without significant adaptation for specific real-time educational uses.
Impact on Educational Practices	Directly applicable for real-time classroom behavior monitoring, enabling immediate pedagogical adjustments based on dynamic student interactions.	Focuses more broadly on technological integration in education, such as hybrid teaching or surveillance, without direct application to real-time behavior analysis and intervention.
Specificity and Novelty	Introduces specific enhancements for detecting nuanced student behaviors, utilizing a targeted approach to improve educational outcomes.	Research generally targets broader AI applications or enhances general model performance, lacking focus on the specific nuances of student behavior in classroom settings.

V. RESULTS AND DISCUSSION

This study endeavors to optimize the SSD algorithm for real-time recognition of student behaviors in classroom settings through the utilization of Mobilenet architecture. The objective of our enhanced algorithm is to augment the effectiveness of educational technology by furnishing precise analyses of student engagement. In this section, we present the outcomes of our experiments and deliberate on their implications for educational practice and future research.

A. Experimental Results

Throughout the experimentation phase, we conducted a comparative analysis of three algorithms: the conventional SSD, the unenhanced Mobilenet-SSD, and our proposed enhanced Mobilenet-SSD. These algorithms were evaluated based on three key metrics: training complexity, detection accuracy, and velocity, to ensure equitable comparisons under consistent conditions.

1) Training complexity: The enhanced Mobilenet-SSD algorithm exhibited a more rapid reduction in loss values during the training process compared to both the conventional SSD and Mobilenet-SSD algorithms. Notably, the enhanced algorithm achieved a loss value of 0.5 in approximately six days, while the conventional SSD algorithm required approximately eight days to achieve a similar reduction.

2) Detection accuracy: Utilizing a bespoke dataset designed to simulate real-world classroom scenarios, we observed that the conventional SSD algorithm attained an accuracy of 84%, the Mobilenet-SSD algorithm achieved 76.14%, and our enhanced Mobilenet-SSD algorithm surpassed both, attaining an accuracy of 85.21%. This enhancement in accuracy is pivotal for the precise identification of student behaviors within classroom environments.

3) Detection velocity: The enhanced Mobilenet-SSD algorithm demonstrated superior detection speed compared to both the conventional SSD and Mobilenet-SSD algorithms. It efficiently processed and identified student behaviors within 21 units, whereas the conventional SSD and Mobilenet-SSD algorithms required 27.1 and 22 units, respectively.

B. Discussion

Our experimentation underscores the efficacy of the enhanced Mobilenet-SSD algorithm in accurately and expeditiously detecting student behaviors within classroom settings. By using Mobilenet's lightweight architecture and optimizing the SSD algorithm, we achieved significant enhancements in both detection accuracy and speed. Several factors contribute to this improvement. Primarily, the integration of Mobilenet architecture alleviated parameter load, thereby expediting training and enhancing efficiency. Moreover, the feature fusion technique bolstered the algorithm's capability to detect small targets, such as writing and mobile phone usage, which are prevalent behaviors in classroom settings. Additionally, the adoption of the RMSProp optimization algorithm further refined detection outcomes by enhancing the accuracy and efficiency of image feature detection. This optimization strategy, coupled with Mobilenet's lightweight design, surpassed traditional SSD methodologies.

C. Implications for Educational Practice

The findings of this study bear substantial implications for educational practice, particularly in the domains of classroom management and student engagement. The enhanced Mobilenet-SSD algorithm equips educators with a potent tool for real-time monitoring of student behaviors, facilitating prompt interventions and personalized instructional strategies. The accurate identification of behaviors such as listening, raising hands, writing, sleeping, and mobile phone usage furnishes educators with valuable insights into classroom dynamics, enabling tailored teaching methodologies. This real-time feedback mechanism fosters student engagement and enhances learning outcomes by addressing individual needs. Furthermore, the algorithm's efficiency facilitates seamless integration into existing educational technologies, enabling scalable deployment across diverse learning environments. Educators can harness this technology to cultivate dynamic and interactive classroom experiences that promote active learning and collaboration among students.

D. Future Research Directions

While this study represents a significant stride in classroom behavior recognition, numerous avenues for future research beckon exploration. Expanding the dataset to encompass a broader spectrum of classroom settings and student demographics could augment the algorithm's generalizability. Furthermore, delving into alternative machine learning frameworks and optimization techniques holds the promise of further augmenting detection accuracy and speed. Longitudinal studies investigating the impact of real-time behavior monitoring on educational outcomes would furnish valuable insights into the algorithm's efficacy in enhancing student engagement and learning. In conclusion, the optimization of the SSD algorithm with Mobilenet architecture presents a promising avenue for enhancing educational technology and classroom management practices. By harnessing advanced machine learning techniques, educators can delve deeper into student behaviors, fostering more inclusive and efficacious learning environments.

VI. Conclusion

In the conventional approach to teaching, it is of utmost importance for teachers and educational institutions to grasp the prevailing conditions within the classroom throughout a designated course through artificial work and thus judge the efficiency of students' lectures, their acceptance degree, and their attendance rate. This research has made significant progress in classroom behavior recognition by improving the Single Shot Detector (SSD) algorithm using Mobilenet architecture for educational purposes. It has set new benchmarks for real-time behavior monitoring, providing valuable insights for educators to enhance classroom dynamics and teaching methods. However, the effectiveness of this refined algorithm is limited by the homogeneous training data, which mainly consists of images from controlled environments. To improve on these results, future studies should expand the dataset to include a wider range of classroom settings and behaviors. Additionally, exploring more advanced machine learning frameworks and conducting longitudinal research would enhance understanding of the impact of real-time monitoring on educational achievements. Ultimately, incorporating sophisticated AI tools in this study not only improves behavior analysis accuracy but also greatly enhances the applications of intelligent educational technology. This study's initial focus involves examining the design process employed in developing the classroom behavior recognition model. Subsequently, the network structure of the conventional SSD model is elucidated, followed by an analysis of its strengths and weaknesses. Later on, the correlation principle of deep separable convolution is elucidated, followed by an introduction to the fundamental network architecture of Mobilenet. Furthermore, a comprehensive analysis is conducted to integrate the distinctive features of each layer within the network. In light of the preparation work mentioned above, the data is processed by optimizing the traditional SSD algorithm after transforming the traditional network. After experimental verification, it is also shown that the optimization algorithm proposed in this paper can detect students' classroom behavior more accurately and quickly. During experimentation, we compared three algorithms: conventional SSD, unenhanced Mobilenet-SSD, and our enhanced Mobilenet-SSD, focusing on training complexity, detection accuracy, and velocity for fair comparisons. The enhanced Mobilenet-SSD algorithm showed faster loss reduction during training than both conventional SSD and Mobilenet-SSD, achieving a loss value of 0.5 in about six days compared to eight days for conventional SSD. Using a custom dataset mimicking real-world classroom scenarios, conventional SSD achieved 84% accuracy, Mobilenet-SSD 76.14%, and our enhanced Mobilenet-SSD outperformed both with 85.21% accuracy, crucial for precise student behavior identification. The enhanced Mobilenet-SSD also exhibited superior detection speed, processing student behaviors within 21 units, while conventional SSD and Mobilenet-SSD required 27.1 and 22 units, respectively. Despite promising results, the study faced limitations due to a homogeneous dataset from controlled environments, potentially impacting the findings' generalizability. Future research should expand the dataset to diverse educational settings and student demographics to test the algorithm's effectiveness. Exploring alternative machine learning frameworks and conducting longitudinal studies would enhance understanding of real-time behavior monitoring's impact on educational outcomes. Pursuing these avenues promises deeper insights and improvements in AI technologies' application in education, enhancing behavior recognition precision and the overall educational experience.

AUTHORSHIP CONTRIBUTION STATEMENT

Dan Liu: Writing-Original draft preparation, Conceptualization, Supervision, Project administration.

Yongqing Cao: Methodology, Software, Validation.

COMPETING OF INTERESTS

The authors declare no competing of interests.

REFERENCES

[1] Huang Y, Yao J, Huang G. Application of intelligent information technology in the reform of hybrid teaching courses in colleges and universities. J Phys Conf Ser, vol. 1852, IOP Publishing; 2021, p. 022065.

[2] Li Y, Qi X, Saudagar AKJ, Badshah AM, Muhammad K, Liu S. Student behavior recognition for interaction detection in the classroom environment. Image Vis Comput 2023;136:104726.

[3] Trabelsi Z, Alnajjar F, Parambil MMA, Gochoo M, Ali L. Real-time attention monitoring system for classroom: A deep learning approach for student’s behavior recognition. Big Data and Cognitive Computing 2023;7:48.

[4] Tran N, Nguyen H, Luong H, Nguyen M, Luong K, Tran H. Recognition of Student Behavior through Actions in the Classroom. IAENG Int J Comput Sci 2023;50.

[5] Park W, Kwon H. Implementing artificial intelligence education for middle school technology education in Republic of Korea. Int J Technol Des Educ 2024;34:109–35.

[6] Sharma SK, Dixit RJ, Rai D, Mall S. Artificial Intelligence and Machine Learning in Smart Education. Infrastructure Possibilities and Human-Centered Approaches With Industry 5.0, IGI Global; 2024, p. 86–106.

[7] Hasnine MN, Nguyen HT, Akçapınar G, Morita R, Ueda H. Classroom Monitoring using Emotional Data 2023.

[8] Shuai Q, Wu X. Object detection system based on SSD algorithm. 2020 international conference on culture-oriented science & technology (ICCST), IEEE; 2020, p. 141–4.

[9] Hu K, Lu F, Lu M, Deng Z, Liu Y. A marine object detection algorithm based on SSD and feature enhancement. Complexity 2020;2020:1–14.

[10] Yan Y. Using the Improved SSD Algorithm to Motion Target Detection and Tracking. Comput Intell Neurosci 2022;2022.

[11] Liu J. The Detection of English Students’ Classroom Learning State in Higher Vocational Colleges Based on Improved SSD Algorithm. International Conference on E-Learning, E-Education, and Online Training, Springer; 2023, p. 96–111.

[12] Zhang W, Xu Q. Optimization of college English classroom teaching efficiency by deep learning SDD algorithm. Comput Intell Neurosci 2022;2022.

[13] Wang S, Xu M, Sun Y, Jiang G, Weng Y, Liu X, et al. Improved single shot detection using DenseNet for tiny target detection. Concurr Comput 2023;35:e7491.

[14] Nandhini TJ, Thinakaran K. Object Detection Algorithm Based on Multi-Scaled Convolutional Neural Networks. 2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP), IEEE; 2023, p. 1–5.

[15] Cheng L, Ji Y, Li C, Liu X, Fang G. Improved SSD network for fast concealed object detection and recognition in passive terahertz security images. Sci Rep 2022;12:12082.

[16] Dai Y. Online English teaching quality assessment based on K-means and improved SSD algorithm. Advances in Multimedia 2022;2022.

[17] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9:1735–80.

[18] Peng D. An English Teaching Pronunciation Detection and Recognition Algorithm Based on Cluster Analysis and Improved SSD. Journal of Electrical and Computer Engineering 2022;2022.

[19] Donahue J, Anne Hendricks L, Guadarrama S, Rohrbach M, Venugopalan S, Saenko K, et al. Long-term recurrent convolutional networks for visual recognition and description. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, p. 2625–34.