

## Research Article

# Neural Network Topology Construction and Classroom Interaction Benchmark Graph Based on Big Data Analysis

Congcong Luan <sup>1,2</sup> and Peng Shang<sup>3</sup>

<sup>1</sup>*School of Marxism, Shandong University, Jinan 250000, China*

<sup>2</sup>*Zibo Normal College, Zibo 255100, China*

<sup>3</sup>*Boshan Middle School, Zibo 255100, China*

Correspondence should be addressed to Congcong Luan; [lcc820125@163.com](mailto:lcc820125@163.com)

Received 27 July 2021; Revised 15 August 2021; Accepted 17 August 2021; Published 20 September 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Congcong Luan and Peng Shang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of artificial intelligence and deep learning in recent years, many universities have put forward the goal of achieving digitalization, intelligent, and education informatization on campus. Throughout the lecture and learning process, the classroom status is an important reference factor to assess students' acceptance of the course and the quality of lectures. However, at present, classroom status analysis is mainly conducted manually, which can distract teachers' attention, so it is of great research significance to find a method that can improve the efficiency of classroom status analysis. In this paper, we choose an offline method to analyze the status of a classroom video recording in terms of students' behavior and attendance in terms of frames, in which student behavior is identified by an improved target detection algorithm and attendance is analyzed by face recognition. By analyzing the structure of the neural network model, an improved neural network model is proposed for its characteristics of a large number of parameters and poor detection of small targets in the basic network. The backbone network is replaced by the improved neural network, and the depth-separable convolutional network is used to reduce the network parameters and increase the computation speed. The information in the deeper feature map is fused upward into the shallow layer to improve the accuracy of small target recognition. Finally, the optimization algorithm is incorporated into the network to optimize the network model and accelerate the model convergence speed. In addition, this paper incorporates the improved behavior recognition method and face recognition method into the system to realize the analysis of the offline classroom status. The system is divided into a teacher side and a management side, where the teacher side is responsible for uploading course recordings and the management side is responsible for randomly analyzing students' status and attendance at any time, and the combination of the two forms a convenient and comprehensive classroom status analysis system platform. Users can upload classroom videos through the instructor interface and can view the classroom status analysis results of a course at any time by searching randomly in the administration. In this paper, the classroom status is mainly judged by the recognition of students' behaviors.

## 1. Introduction

In the era of rapid development of technology, college education is also progressing, mainly in the aspects of campus life and classroom teaching [1]. It mainly includes "Intelligent Library," "Smart Campus Card," "Access Control Management," and "Smart Dormitory" [2]. Previously, you needed a meal card to eat, a library card to borrow books, a bath card to take a bath, and an access card to enter

and leave the dormitory, but the mobile smart card integrates these functions into one campus one-card [3]. Meanwhile, reforms in classroom teaching are also ongoing, and many innovative educational theories have emerged, such as flipped classroom, teaching theory, constructivism theory, and theme-based education theory. These theories have given great impetus to the teaching reform. However, there are still many problems in traditional classroom teaching [4]. At present, classroom attendance is mainly in the form

of roll call, which is fine in small classes, but not in large classes with hundreds of students, where it takes about thirty minutes to roll call everyone [5]. Not only is it time-consuming and labor-intensive, but there will also be people with answers. Another way to do roll call is to sign in via mobile app. Although this way saves time, there will be network lag when there are more people, so students cannot sign in on time; in addition, there are also cases of people taking the cell phone and signing in for them [6].

As the main subject of classroom teaching, students' classroom behavior can reflect the degree of acceptance of knowledge and will also directly affect the quality of teachers' lectures [7]. In traditional classroom teaching, teachers can only understand the status of a class through their own observation, which undoubtedly increases the workload for teachers, and the results obtained can be rather one-sided, so there is an urgent need for methods that can improve the efficiency of classroom status analysis [8]. With the development of artificial intelligence, using it to improve campus life and teaching quality has become an important direction for future research in education. A study of the development of AI reveals that it first appeared in 1943, reached its golden age in 1956-1974, and then experienced two lows and a boom. Nowadays, after a long development, it has become an interdisciplinary discipline, penetrating into all walks of life and even daily life, especially in the field of education, which has been developing rapidly in recent years [9]. On the integration of AI and education, it is pointed out that the research on the application of AI in education is at a historical inflection point, starting from the emergence of research on AI in education in 2017 to the explosive growth of related research from 2018. It is the future development trend to use innovative technologies, reform the methods used in teaching, construct a new education system containing innovative learning and interactive forms of learning, and realize intelligent teaching in accelerating the construction of the talent training model [10]. The target detection and face recognition algorithms used in this paper are part of the artificial intelligence applications. Target detection is used to locate and classify targets, for example, in intelligent driving to determine whether there are obstacles on the road and the exact location of the obstacles. The main purpose of face recognition is to identify who the detected face is and is divided into two forms: 1:1 and 1:N. Among them, 1:1 solves the problem of "whether this person is someone," such as judging whether the ID card is the person in the security check [11]. And 1:N solves the problem of "who is this person" by finding the identity of this person in a large amount of face data based on the image provided. Using the target detection method to identify student behavior, we can quickly count the number of people in the classroom for each behavior, which is convenient and accurate compared to manual methods. Using face recognition in classroom attendance, it is possible to automatically identify people who have arrived by simply entering the class image, which is fast and avoids phenomena such as proxy signing [12]. Therefore, using deep learning to analyze classroom status has practical research value and significance [13].

The training set is put into the behavior recognition network model for training to get the initial model, and the validation set is used to validate the model and adjust the network model parameters according to the validation results. The test data is put into the model to get the results, and then the results are analyzed to determine if they are the same as what is expected. Through the analysis of the previous contents, we learned that the relevant techniques and theories have been developed more and are more advanced, but not many of them have been applied to students' classroom states. Therefore, in this paper, we review the literature on target detection technology and face recognition technology, study the application methods of both in other fields, and improve or migrate the two algorithms according to the characteristics of students' classroom status to make them suitable for this study. In this paper, we choose a model combining the MTCNN face detection algorithm in pyramid mode and InsightFace face comparison algorithm for face recognition, which can meet the effect of real-time recognition and has a high accuracy rate in practice. The system is designed and implemented using the Flask+Vue framework, divided into two main parts: the teacher side and the management side. The teacher side is the front-end of the system responsible for uploading classroom video recordings to the database and viewing student status analysis and attendance results, while the management side is the back-end of the system responsible for personal management, comprehensive query, status statistics, attendance management, teaching resources, and management of basic resources.

## 2. Related Work

The traditional algorithm for detection of targets can be summarized in three main steps: first, selecting candidate regions in the image; second, extracting the features in the candidate regions where they are located; and finally, recognizing them using a classifier in the last step [14]. The Harr-like features proposed by Jakonen et al. to describe the face, also known as Harr features, can be divided into linear features, edge features, point features, and diagonal features; they proposed a method to obtain the features of the directional gradient histogram, by which the features are obtained. The way to get features is to count and calculate the gradient histogram for a certain range of the image [15]. The method of manually designed features has some drawbacks, such as the problem that the image features are not obvious and difficult to set, the designed features cannot be used under certain conditions, and the robustness is not good [16]. The AdaBoost method was proposed by Gardner, who automatically selected some weak classifiers with low recognition accuracy and combined them to increase the judgment [17]. Support vector machines achieve binary classification of data by supervised learning and have higher accuracy on linearly separable datasets, often by introducing kernel functions to map a low latitude to a high latitude to make the data linearly separable.

Decision tree is an inductive learning method based on examples, which distills a tree classification structure in

unordered data and then makes classification judgments. Random forest consists of many decision trees and classifiers, and the plurality of all the trees contained determines their output results, and the method will improve the classification accuracy [18]. Deep learning-based target detection is gradually becoming mainstream, and its detection accuracy is much higher than the traditional one. Researchers combined the RCNN algorithm and the SPP-net algorithm to propose the Fast R-CNN algorithm, which incorporates object recognition and position correction into a single network for training to reduce memory usage. The Faster R-CNN algorithm is based on the Fast R-CNN with the addition of a region generating network (RPN), which increases the speed of the algorithm [19]. The above-mentioned algorithm is called a two-stage algorithm, which classifies the candidate frames first and then is not fast enough although it has high accuracy [20]. Therefore, researchers first proposed the one-stage YOLO algorithm, which combines the generation of candidate frames and classification together, and the recognition accuracy is worse than that of Faster R-CNN but faster. This method is comparable to Faster R-CNN in terms of accuracy but much faster than it [21].

In the field of target detection, researchers have conducted a lot of research and achieved good results. The researchers improved the algorithm of YOLO V3 [22] by increasing the size of the feature map directly, instead of increasing a portion of it and then combining it with the residual blocks [23] to improve the detection of small targets. The researchers proposed a small target detection algorithm based on Faster-RCNN combined with multiscale feature fusion [24] and online hard case mining for airport scenes to improve the accuracy of small target object detection in airport scenes. Computer vision [25–27] development goal is to make it gradually close to the human eye function or even better than the human eye to complete some difficult tasks. For example, for the most applied face recognition now, the technology not only has the identity recognition function outside like human eyes but also is much faster, which can quickly compare from a large amount of data to find the target face. The development and application of computer vision and surveillance facilities are inseparable, and the combination of the two has been widely used in intelligent surveillance, real-time patient monitoring [28], virtual reality, intelligent robotics, etc. Pierce divided the computer vision based on video surveillance into three directions in his paper: behavior recognition [29] and analysis, tracking detection, and motion target detection, and proposed that behavior recognition analysis is the main direction of future development. A search on the Internet for the keyword “behavior recognition” yields 8112 results, which shows that the research on it is very high. By reading the relevant literature, it is concluded that the development of behavior recognition can be divided into two major directions: behavior recognition with artificially selected features and behavior recognition based on deep learning [30–32], and each direction can be divided into more detailed categories according to different research methods.

### 3. Neural Network Recognition Model for Classroom Interaction Benchmark Map

*3.1. Construction of Classroom Behavior Recognition Model.* The construction of the model mainly includes determining the network structure, preparing training data, and training and testing the model, which leads to the design flow of the behavior recognition model as shown in Figure 1.

The first step was to prepare images of student behaviors. We collected 2500 images of each of the five student behaviors: raising hands, sitting, writing, sleeping, and playing with a cell phone, and 500 images of each behavior. The second step was to build a behavior recognition database. The collected 2500 images were preprocessed and labeled, and the images were proportionally divided into three parts: training set, test set, and validation set. Finally, the model is trained and tested. The training set is put into the behavior recognition network model for training to get the initial model, and the validation set is used to validate the model and adjust the network model parameters according to the validation results. The test data is put into the model to get the results, and then, the results are analyzed to determine if they are the same as what is expected. We decide whether to continue training the model based on the comparison results. We save the behavior recognition models with better recognition results for subsequent classroom behavior recognition.

*3.2. SSD Target Detection Algorithm.* The target detection algorithm used in this paper is based on the improvement of the traditional SSD target detection algorithm, so the structure of the traditional SSD algorithm model and the principle of the algorithm are briefly explained before introducing the improved algorithm. SSD algorithm, a regression-based target detection model, was proposed in 2016. Depending on the input image size, SSD is divided into SSD300 and SSD512, and SSD300 is used in this paper. Its network structure is divided into two parts: first is the main part of the network, which is generally called the base network, derived from some subtype networks; second is the convolutional network added subsequently, which is used to help the previous network to further extract image features. The last fully connected layer of VGG16 is removed, and only the previous convolutional network is used, and the two new convolutional layers, named Conv6 and Conv7, are placed in the place of the just-deleted fully connected layer, and finally, eight decreasing convolutional layers are added at the end, followed by the classification layer and the nonmaximum suppression layer, as shown in Figure 2.

According to the previous section, SSD and YOLO are both one-stage-type target detection algorithms, but they are different in terms of feature extraction. The early YOLO algorithm only extracts the information of the top-level features by convolution operation, which is semantic but may lose the information of small targets, so as mentioned above, the early YOLO algorithm is fast but the detection rate of small targets is not high. The SSD algorithm uses multiple scales of feature maps for detection and adds gradually decreasing convolutional layers after the modified VGG16 base network and then selects six layers for prediction from

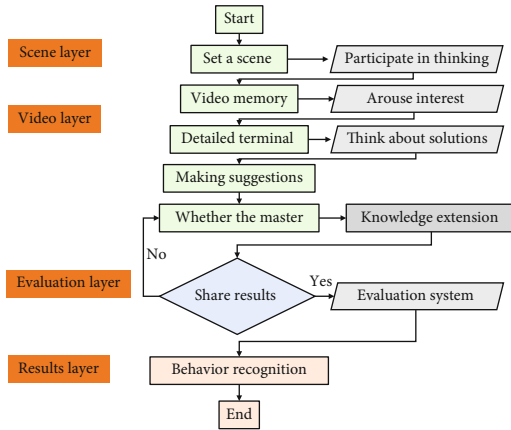


FIGURE 1: Classroom behavior recognition process.

all the layers, which are Conv4\_3, Conv7, Conv8\_2, Conv9\_2, Conv10\_2, and Conv11\_2, and their sizes are from front to back. The feature map size is gradually effective, where the larger ones are used to identify the smaller objects and the smaller ones are used to identify the larger objects. The SSD algorithm mimics the Anchor mechanism in the Faster R-CNN, but the two are very different in the application of Anchor. Faster R-CNN is a classical two-stage target detection algorithm consisting of Fast R-CNN and RPN, where the core of RPN is the Anchor mechanism responsible for generating target regions. The detection is done in two parts: first, the most likely regions are identified by a network dedicated to selecting regions, and then, the objects in the regions are classified by a target classification network. This completes the category prediction. In addition, unlike Faster R-CNN, which only uses Anchor in the last layer to generate candidate regions, SSD uses the Anchor mechanism in all the six different size feature maps mentioned above to achieve the purpose of multiscale detection.

**3.3. Topology Improvement of Neural Network.** From the introduction of the principle of the SSD algorithm above, it can be seen that the SSD algorithm mainly extracts image features through the underlying network and additional convolutional layers and then selects some feature layers for the target detection work. Although this algorithm has achieved good results in the field of target detection, there are still some shortcomings. The traditional SSD algorithm uses a modified VGG16 as the base network, which has a good classification effect but has a large number of parameters [33], i.e., 14,122,995 parameters excluding the final fully connected layer, and about 4/5 of the training time is spent on the base network, so it is difficult to train and requires a high computer configuration, and the real-time performance is not good enough. In addition, the network structure of the SSD algorithm shows that the detection of small targets is done at the shallow level of the feature map, but this level contains less information about the features, so the detection effect is not good enough. In this section, we propose the following improvement strategies for both the basic network and the small target detection, replacing VGG16 with a lightweight network to improve the detection

speed by reducing the number of parameters, and fusing the high-level semantics to the lower level to improve the small target detection. The improvement principle and the process are described in detail in the following. The goal of the basic network improvement is to replace the original backbone network VGG16 with a lightweight network. It uses deep separable convolution instead of normal convolution to reduce the number of parameters, and Mobilenet has only 4.2 million parameters compared with the 133 million parameters in VGG16. The results of both tests on the ImageNet dataset show that Mobilenet is much faster and the accuracy is only 0.9 percentage points lower than that of VGG16. Therefore, this paper uses the original Mobilenet as the base network for SSD with some modifications. The Mobilenet improvement and base network replacement process are described in detail as follows.

The topological network is faster and less computationally intensive than, for example, VGG16 because it differs in two ways: firstly, the network is composed with depth-separable convolution, and on the other hand, the width and resolution coefficients are also used. The main part is the depth-separable convolution, which completes a convolution operation in two parts: depth convolution and point convolution. The Mobilenet network structure has 28 layers if both are considered as two layers and 14 layers if they are considered as one layer. As mentioned above, the depth-separable convolution means that the convolution operation is implemented in two steps. When the image is input to the network, it needs to go through the deep convolution operation to get some maps containing the feature information, then the BN and ReLU operation to get some other feature map information after the feature map obtained above, and then the BN and ReLU operation again to get the result. The process is shown in Figure 3, with depth convolution on the top and point convolution on the bottom.

In order to reduce the network parameters, in addition to the depth-separable convolution, the width factor  $\alpha$  and the resolution factor  $\rho$  are used, with values between 0 and 1. The most common values of  $\alpha$  are 1, 0.75, 0.5, and 0.25.  $\alpha$  serves to reduce the number of channels, for example, for an input channel with a value of  $R$ , when added, it becomes  $\alpha R$ , which reduces the computational effort by about  $\alpha 2$ . Another factor that affects the amount of computation is the resolution, so  $\rho$  is used to reduce the resolution of the image, and the use of this factor reduces the amount of computation of pixel values by  $\rho 2$  similar to the reduction of  $\alpha$ .

## 4. Results and Analysis

Big data provides data support for exploring personalized group teaching models, which helps to effectively promote the transformation of group teaching to individualized student learning. Teachers can use high-end big data analysis technology, which can focus on the overall microlearning performance of students at each stage in real time, such as student response scores for each multiple-choice question and long-term student learning performance in school. The big data learning platform will provide school teachers with timely, authentic, and unique teaching information in order



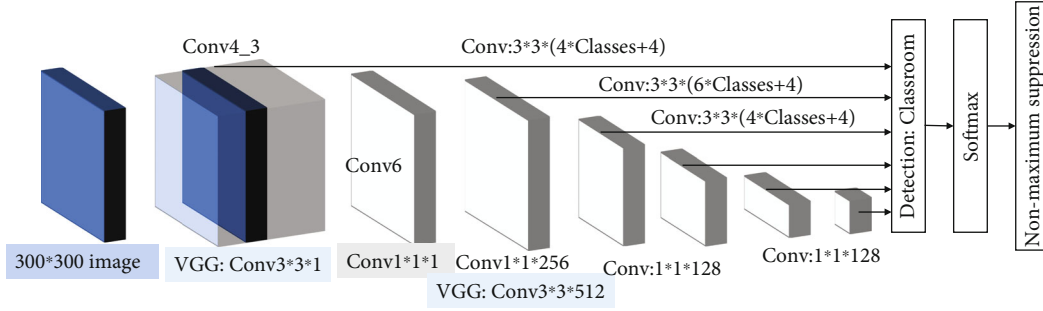


FIGURE 2: SSD network structure.

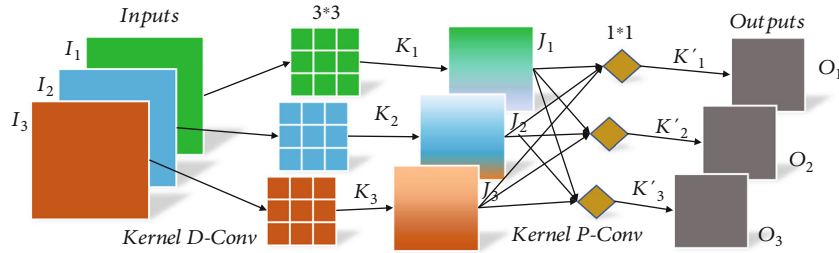


FIGURE 3: Deeply separable convolution operations.

to help teachers tailor their teaching in the classroom management process—which stages of the learning management process in the teacher’s classroom should students pay special attention to when reviewing the basics section, which stages should students pay special attention to when incorporating the practical class, at which stage should students pay special attention for the review of the basics, at which stage should they pay special attention for the integration of the contents of practical classes, at which stage should they pay attention for the strengthening of comprehensive exercises, and at which stage can they concentrate on reading the books recommended by the teacher, etc. In addition, after successfully completing the mathematics assignments assigned by the school teachers, students may also use the Smart Learning Assessment System to continuously strengthen their independent learning. The intelligent system [34, 35] is to recommend mathematics homework; if some students can answer all the recommended questions of a certain type several times correctly, they can naturally skip similar recommended questions; on the contrary, the effect is more intensive, so that not only can it greatly improve the efficiency of the candidates in learning but also it can greatly reduce the burden of some students in later learning.

Through the data collection of daily assignments and exams, we can understand the common problems and individual problems of the class, and the teacher can explore the shortcomings of the students through the data analysis. In the process of data interpretation, the teacher can compare the average score of each class, analyze the gap of each question, and find the common problems and individual problems of the class, and the teacher can strengthen the training of students for the common problems. Individual problems can be instructed individually. There may be some

obstacles between the information of the questions in the test paper or homework and the students’ goal achievement, and these obstacles can be found through the data analysis of the big data products of the extreme class, which is difficult to dig based on manual collection. The high-frequency errors and error-prone questions of the student population are also a reflection of their lack of knowledge and problem-solving skills. As academic data is collected, each student’s level in the academic diagnostic evaluation is automatically divided. Combining the students’ levels and scores, we can identify common problems and focus on cultivating excellence and transforming late-comers. With the collection of academic data, each student’s “Classmates” account will accumulate the wrong questions in the subject, forming an electronic diagnostic academic file for each student, which can be revised on cell phones and tablets with “Classmates,” and can also be exported and printed into a book. This is an invaluable resource for students and provides a solid foundation for adaptive learning research and personalized tutoring.

In this paper, we evaluate the model in terms of single-frame image detection time and mean average precision ( $P$ ) of image detection. AP is the average of all class  $P$  values. AP is the area under the line of the curve consisting of precision and recall.

The accuracy formula is as follows:

$$P = \frac{Tp + Fp}{(Tp^2 + Fp^2)^{1/2}}, \quad (1)$$

where  $Tp$  means the number of samples for which the classifier classifies the target as a positive sample and is

actually a positive sample as well. Fp refers to the number of samples that the classifier considers as positive samples but is actually negative samples.

The whole equation represents the proportion of positive samples that the classifier considers as positive samples and is indeed a positive sample itself to the overall positive samples identified by the classifier, reflecting the checking function of the model.

The formula for the recall rate is as follows:

$$\text{Recall} = \frac{\text{Tp}^{1/2}}{(\text{Tp}^2 + \text{Fp}^2)^{1/2}}, \quad (2)$$

where TN refers to the number of samples for which the classifier considers the target as negative samples but is actually positive; the whole equation indicates the proportion of samples that the classifier considers positive classes and is indeed positive classes to all positive classes, reflecting the check-all function of the model.

The squared error criterion is generally used, and the relevant definition is as follows:

$$E(n) = \iint (\lambda t)^{n+1} e^{\lambda t}. \quad (3)$$

The algorithm uses probability-based distances as a measure function:

$$d(k, n) = (\lambda t)^{n+1} e^{\lambda t} + \xi_i, \quad (4)$$

$$\zeta_n = H_{N,v} \left( \sum_{i=1}^k \sum_{P \in Q} (\lambda t)^{n+1} e^{\lambda t} + Q_n \right),$$

$$\tilde{E}(n) = \sum_{i=1}^k \sum_{P \in Q} (\lambda t)^{n+1} e^{\lambda t} \log_n \frac{N_v}{N_n}.$$

The traditional SSD algorithm, Mobilenet-SSD, and Mobilenet-SSD with feature fusion are trained under the same experimental environment and parameters, and then, the three algorithms are compared by a test set. The test environment is shown in Figure 4, and the data used for the test are the homemade datasets in this paper. Figure 4 shows the average accuracy and detection speed (detection time per image frame) of classroom behavior detection obtained using different models on the test set.

It can be clearly seen from Figure 4, we can see that in the student classroom behavior recognition test, the Mobilenet-SSD with feature fusion improves the detection speed by 2 frames per second compared with the traditional SSD algorithm, and the average accuracy of detection reaches 83.08%. Compared with the Mobilenet-SSD model without feature fusion, the speed is reduced by 2.5% because the fusion increases the network parameters, but the accuracy is improved by 6.94%. After the above analysis, this algorithm has good performance compared with the other two algorithms in terms of detection speed and recognition accuracy. Comparing the loss function change curves during the training process, we can judge the training difficulty of

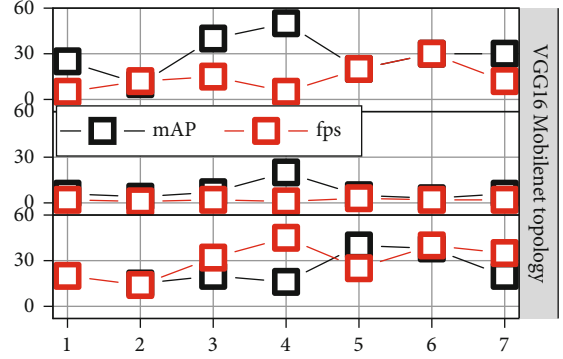


FIGURE 4: Comparison result of recognition effect of different models.

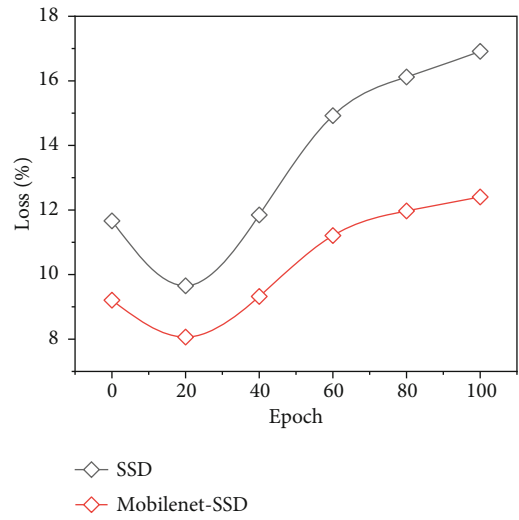


FIGURE 5: Loss variation graph.

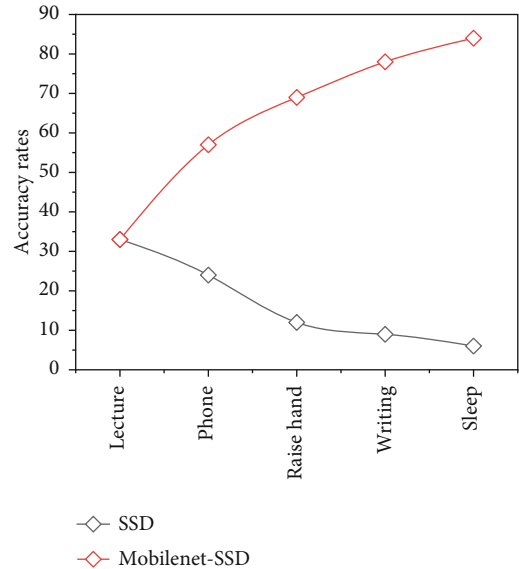


FIGURE 6: Comparison of the accuracy rates of the five classroom behavioral tests.

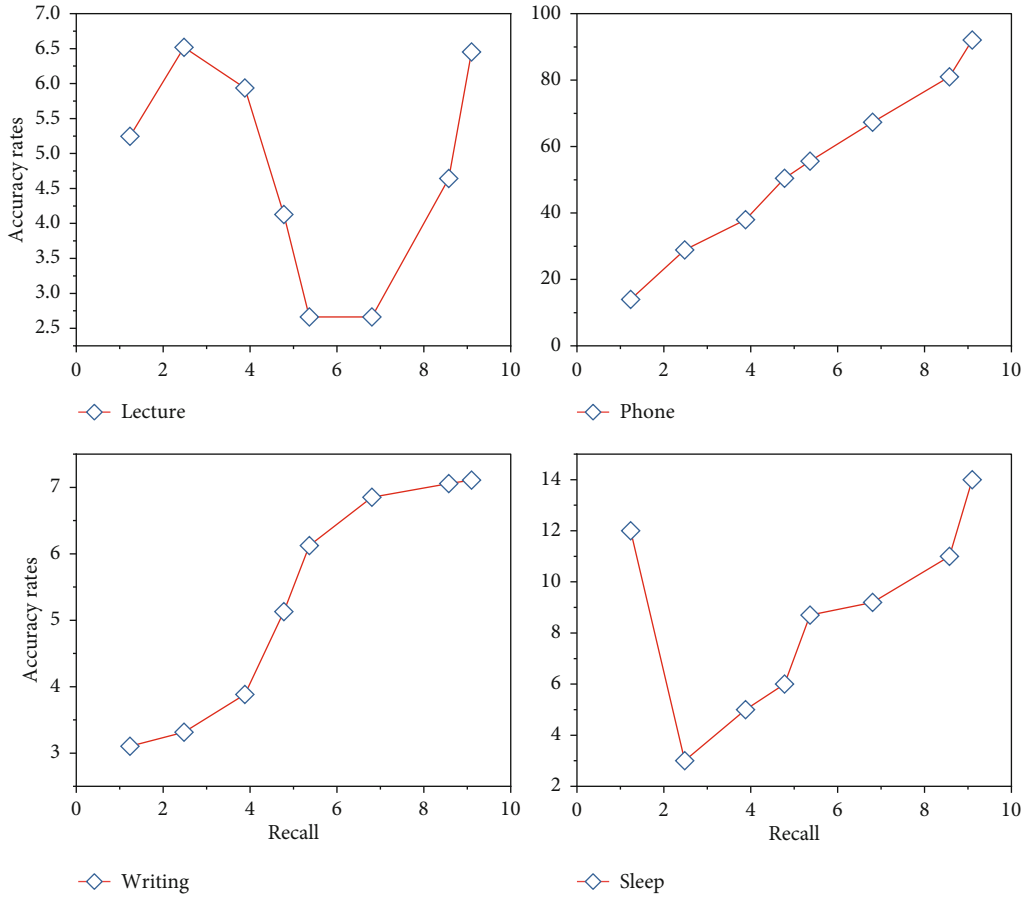


FIGURE 7: Recognition accuracy graph.

the model. The loss function curves of Mobilenet-SSD and SSD models with the same parameters and epoch of 100 for 50,000 iterations are shown in Figure 5.

It can be seen that the loss values of both models have been decreasing, which means that both models are reasonable. It took about 6 days for the loss to fall below 0.5 during training and about 8 days for the original SSD model, and it can be seen from the graph that the loss value of this model decreases faster than that of the traditional SSD model. The accuracy (AP) of each action is shown in Figure 6, which is obtained by using the Mobilenet-SSD model after SSD and feature fusion to detect five actions of students in the test set: listening, raising hands, writing, sleeping, and playing with cell phones.

From Figure 6, it is concluded that the Mobilenet-SSD algorithm after feature fusion in this paper has improved the detection of small targets in all five actions compared with the original SSD algorithm, among which writing has improved the most by 3.03%, indicating that the model in this paper has improved in small target recognition. The results of the Mobilenet-SSD recognition of feature integration showed that among the five actions, the detection accuracy of listening action was the highest and hand raising was the second highest, while writing and playing with cell phone were the lowest. The reason for this result is that the two actions are more likely to be obscured than the other

actions, especially in the recognition process, which can easily confuse other hand movements, and thus, the calibration effect is not as good as the other three actions. In order to show more intuitively the recognition accuracy of the model on the five actions, a line graph is shown in Figure 7, where the shaded area is the accuracy rate.

This paper first analyzes the classroom behavior recognition model design process and then explains how the database is constructed, including data acquisition, image preprocessing, and dataset labeling. Then, the network structure of the traditional SSD model is described, and its advantages and disadvantages are analyzed. After that, the principles related to depth-separable convolution are explained, based on which the basic network structure of Mobilenet is introduced, and the methods that are more frequently used in fusing the features of each layer of the network are analyzed. Based on the above preparatory work, an improved strategy is proposed to change the basic network to Mobilenet and to use the add method to complete the fusion of features, so that the model can improve the effect of recognizing small objects compared with the original SSD method, and the recognition speed is similar to the previous one, as shown in Figure 8.

Only the Gs plays a role, which is equivalent to a Gaussian filter that mainly acts as an image smoother. In an image, there are obvious color or light/dark shifts in the critical areas, which are reflected at the pixel level by the large

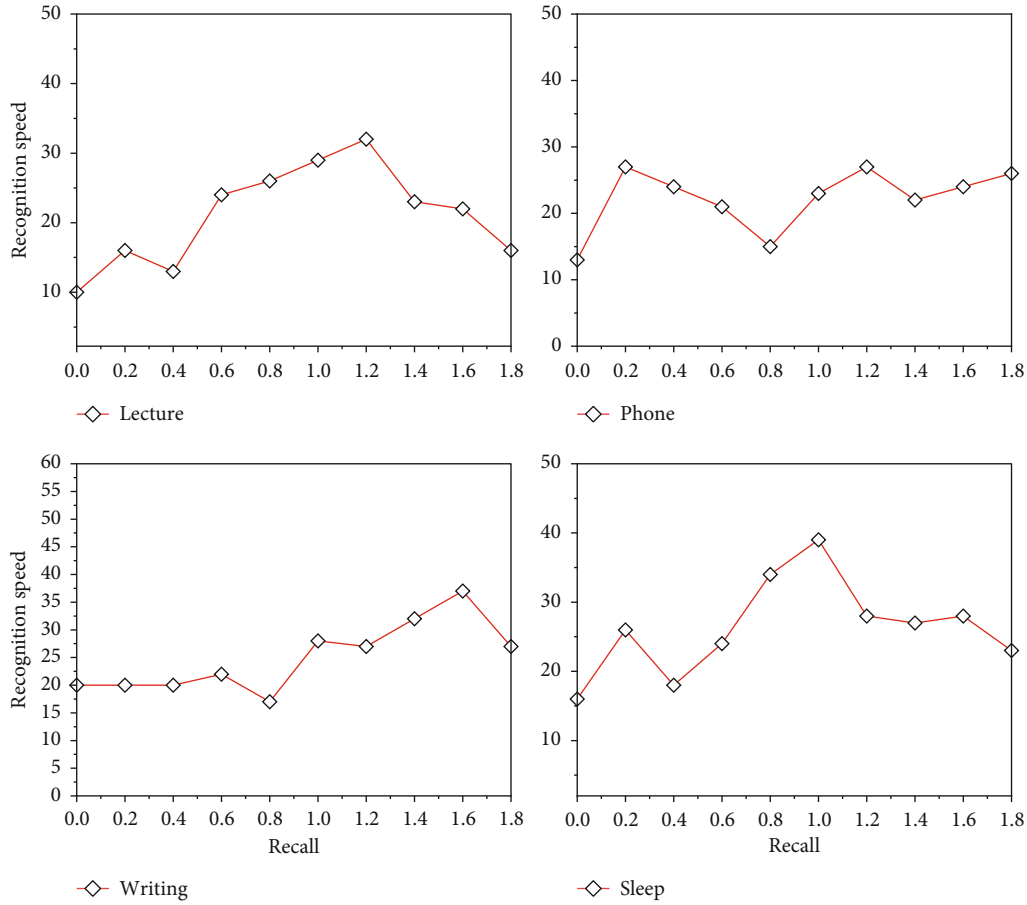


FIGURE 8: Recognition speed comparison.

difference in pixel values between the two sides. In this case, the Gr value is close to 0, and the whole filter result is 0. At this point in the whole filtering process, the value there does not affect the overall output value, which plays a role in protecting the edge effect. Image target enhancement [36] is also called sharpening, using methods such as Sobel (Sobel) processing method, USM algorithm, Laplace (Laplace) processing method, the first-order form of Prewitt processing method, and the canny processing method involving multiple orders; by comparison, this paper uses USM (Unsharpen Mask) sharpening enhancement algorithm for target enhancement. For the management, it is important to keep abreast of the teaching and learning process. For the management, it is more timely to grasp the teaching situation. What teachers do is conscientious work, and real data makes teachers' work results more tangible and transparent. In the process of teaching management implementation follow-up, you can clearly understand who is doing well and who is not doing well, so as to praise teachers for excellence in time and remind teachers for deficiencies in time. The basic theoretical principle of teaching is to teach according to the material, and the basic premise of teaching according to the material is to read and understand students and the important way to read and understand students which is also the academic diagnosis of teachers and students. Only through the academic diagnosis of the candidates is it possible to really read

a student, and by reading a student is it possible to really implement the teaching according to their abilities. Due to the lack of analysis of the teachers' individual teaching tracking survey, it is not conducive to the effective organization of individualized teaching and differentiated teaching of students, which ultimately does not achieve the effect of teaching according to their abilities. By continuously carrying out this personalized teaching, differentiated teaching, it may not find the growth point on the quality of teaching. To improve the examination promotion rate, you can only do personalized and differentiated teaching without extending the time, and to achieve personalized and differentiated teaching, use the extreme class big data.

## 5. Conclusion

In this paper, by changing the base network in the original SSD from VGG16 to the improved Mobilenet network, and using the add feature fusion method for the replaced network, the base network parameters are reduced while deep information is incorporated in the shallow layer, thus improving the effect and speed of small target detection. Then, the model is trained and the trained model is used to identify the student behaviors and get the distribution of the number of students in each of the five behaviors in a class, and then, we get the percentage of the number of



students in the five states of serious, good, fair, poor, and other as set in the thesis to complete the analysis of the student states of the class. This is combined with the thesis algorithm model to design and implement a college classroom status analysis system. Through daily homework and test data collection, we can understand common and individual problems in the classroom, and teachers can explore students' deficiencies through data analysis. In the process of data interpretation, teachers can compare the average scores of each class, analyze the gaps in each question, and find common and individual problems in the class. Teachers can strengthen the training of students for common problems. The web-oriented classroom status analysis system is designed and implemented by using the Flask+Vue framework and combining the behavioral recognition and face recognition algorithm parts. Users can upload classroom videos through the instructor interface and can view the classroom status analysis results of a course at any time by searching randomly in the administration. In this paper, the classroom status is mainly judged by the recognition of students' behaviors. Although the model recognition effect is ok, the reference factors are a bit one-sided, and the classroom status analysis can be done together with the recognition judgment of facial expressions in the future. Secondly, there is room for expansion of this system.

At present, it only analyzes the overall situation of a course during the class, and in the future, it can add the status analysis function for each student and generate the report of students' classroom status analysis according to the statistical results.

### Data Availability

The data used to support the findings of this study are included within the article.

### Conflicts of Interest

All the authors do not have any possible conflicts of interest.

### Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant Nos. 61675164 and 61827827).

### References

- [1] C. Chang and H.-C. K. Lin, "Classroom interaction and learning anxiety in the IRs-integrated flipped language classrooms," *The Asia-Pacific Education Researcher*, vol. 28, no. 3, pp. 193–201, 2019.
- [2] S. K. Ertesvåg, P. Sammons, and U. Blossing, "Integrating data in a complex mixed-methods classroom interaction study," *British Educational Research Journal*, vol. 47, no. 3, pp. 654–673, 2021.
- [3] M. Reed, "Classroom interaction and the development of empowerment," *Changing English*, vol. 25, no. 2, pp. 115–134, 2018.
- [4] A. M. Blaine, "Interaction and presence in the virtual classroom: an analysis of the perceptions of students and teachers in online and blended advanced placement courses," *Computers in Education*, vol. 132, no. 1, pp. 31–43, 2019.
- [5] O. Metatla, A. Thieme, E. Brulé, C. Bennett, M. Serrano, and C. Jouffrais, "Toward classroom experiences inclusive of students with disabilities," *Interactions*, vol. 26, no. 1, pp. 40–45, 2018.
- [6] A. Ogan, "Reframing classroom sensing," *Interactions*, vol. 26, no. 6, pp. 26–32, 2019.
- [7] V. Pavón Vázquez and M. C. Ramos Ordóñez, "Describing the use of the L1 in Clil: an analysis of L1 communication strategies in classroom interaction," *International Journal of Bilingual Education and Bilingualism*, vol. 22, no. 1, pp. 35–48, 2019.
- [8] K. Solheim, "Teachers' aspirations to improve their classroom interaction," *International Journal of Learning Teaching and Educational Research*, vol. 18, no. 6, pp. 147–169, 2019.
- [9] C. Bokhove, "Exploring classroom interaction with dynamic social network analysis," *International Journal of Research & Method in Education*, vol. 41, no. 1, pp. 17–37, 2018.
- [10] K. Solheim, S. K. Ertesvåg, and G. Dalhaug Berg, "How teachers can improve their classroom interaction with students: new findings from teachers themselves," *Journal of Educational Change*, vol. 19, no. 4, pp. 511–538, 2018.
- [11] J. Batlle and M. Deal, "Teacher epistemic stance as a trouble in foreign language classroom interaction," *Journal of Pragmatics*, vol. 176, pp. 15–25, 2021.
- [12] H. Gresch and M. Martens, "Teleology as a tacit dimension of teaching and learning evolution: a sociological approach to classroom interaction in science education," *Journal of Research in Science Teaching*, vol. 56, no. 3, pp. 243–269, 2019.
- [13] U. Karvonen, L. Tainio, and S. Routarinne, "Uncovering the pedagogical potential of texts: curriculum materials in classroom interaction in first language and literature education," *Learning Culture and Social Interaction*, vol. 17, pp. 38–55, 2018.
- [14] A.-M. Vogler, S. Prediger, U. Quasthoff, and V. Heller, "Students' and teachers' focus of attention in classroom interaction — subtle sources for the reproduction of social disparities," *Mathematics Education Research Journal*, vol. 30, no. 3, pp. 299–323, 2018.
- [15] T. Jakonen, "Retrospective orientation to learning activities and achievements as a resource in classroom interaction," *The Modern Language Journal*, vol. 102, no. 4, pp. 758–774, 2018.
- [16] P. L. Hartwick, "Investigating research approaches: classroom-based interaction studies in physical and virtual contexts," *ReCALL*, vol. 30, no. 2, pp. 161–176, 2018.
- [17] R. Gardner, "Classroom interaction research: the state of the art," *Research on Language and Social Interaction*, vol. 52, no. 3, pp. 212–226, 2019.
- [18] S. K. Ertesvåg and T. Havik, "Students' proactive aggressiveness, mental health problems and perceived classroom interaction," *Scandinavian Journal of Educational Research*, vol. 65, no. 1, pp. 1–20, 2021.
- [19] M. S. Lerang, S. K. Ertesvåg, and T. Havik, "Perceived classroom interaction, goal orientation and their association with social and academic learning outcomes," *Scandinavian Journal of Educational Research*, vol. 63, no. 6, pp. 913–934, 2019.
- [20] T. R. D. Roh and Y. A. Lee, "Teacher repetition as an instructional resource for classroom interaction: three pedagogical actions in kindergartens in an Efl context," *System*, vol. 74, pp. 121–137, 2018.

- [21] A. Llinares and N. Evnitskaya, "Classroom interaction in CLIL programs: offering opportunities or fostering inequalities?," *Tesol Quarterly*, vol. 55, no. 2, pp. 366–397, 2021.
- [22] Z. Huang, P. Zhang, R. Liu, and D. Li, "Immature apple detection method based on improved Yolov3," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 9–13, 2021.
- [23] W. Cai, Z. Wei, R. Liu, Y. Zhuang, Y. Wang, and X. Ning, "Remote sensing image recognition based on multi-attention residual fusion networks," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 1–8, 2021.
- [24] C. Gao, S. Ye, H. Tian, and Y. Yan, "Multi-scale single-stage pose detection with adaptive sample training in the classroom scene," *Knowledge-Based Systems*, vol. 222, p. 107008, 2021.
- [25] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020.
- [26] Z. Chu, M. Hu, and X. Chen, "Robotic grasp detection using a novel two-stage approach," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 19–29, 2021.
- [27] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local–global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.
- [28] L. Yu, S. Tao, W. Gao, and L. Yu, "Self-monitoring method for improving health-related quality of life: data acquisition, monitoring, and analysis of vital signs and diet," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 24–31, 2021.
- [29] X. Ning, K. Gong, W. Li, L. Zhang, X. Bai, and S. Tian, "Feature refinement and filter network for person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, In press.
- [30] W. Chu, *Studies on the effects of wiring density on chip package interaction and design optimization with machine learning*, [Ph.D. thesis], The University of Texas at Austin, 2021.
- [31] Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, "Deep learning based cell classification in imaging flow cytometer," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 18–27, 2021.
- [32] W. Cai, Y. Song, and Z. Wei, "Multimodal data guided spatial feature fusion and grouping strategy for E-commerce commodity demand forecasting," *Mobile Information Systems*, vol. 2021, Article ID 5568208, 14 pages, 2021.
- [33] M. Miao, W. Cai, and X. Li, "Parameter estimation of gamma-gamma fading with generalized pointing errors in FSO systems," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1301878, 14 pages, 2021.
- [34] Y. Jiang, X. Gu, D. Wu et al., "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2021.
- [35] A. Soller, "Supporting social interaction in an intelligent collaborative learning system," *International Journal of Artificial Intelligence in Education*, vol. 12, no. 1, pp. 40–62, 2001.
- [36] L. Zhang, X. Wang, X. Dong, L. Sun, W. Cai, and X. Ning, "Finger vein image enhancement based on guided tri-Gaussian filters," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 17–23, 2021.