

Retraction

Retracted: System Analysis of the Learning Behavior Recognition System for Students in a Law Classroom: Based on the Improved SSD Behavior Recognition Algorithm

Scientific Programming

Received 18 July 2023; Accepted 18 July 2023; Published 19 July 2023

Copyright © 2023 Scientific Programming. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their

agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Q. Guo, "System Analysis of the Learning Behavior Recognition System for Students in a Law Classroom: Based on the Improved SSD Behavior Recognition Algorithm," *Scientific Programming*, vol. 2022, Article ID 3525266, 11 pages, 2022.

Research Article

System Analysis of the Learning Behavior Recognition System for Students in a Law Classroom: Based on the Improved SSD Behavior Recognition Algorithm

Qingmei Guo 

Shengda University of Economics, Business & Management, Zhengzhou 451191, Henan, China

Correspondence should be addressed to Qingmei Guo; 100082@shengda.edu.cn

Received 19 July 2022; Revised 10 August 2022; Accepted 7 September 2022; Published 30 September 2022

Academic Editor: Lianhui Li

Copyright © 2022 Qingmei Guo. 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.

Classroom teaching activities have always been the focus of research in the field of pedagogy. The main body of classroom teaching activities is students, and students' classroom behavior status can reflect classroom efficiency to a certain extent, making it an important reference index for classroom quality assessment. With the rapid development of artificial intelligence, school education is gradually becoming more intelligent. At present, most of the classrooms are equipped with video equipment. These videos record the real behavior status of the students in the classroom. For example, by analyzing the data, combining artificial intelligence, deep learning, and other related technologies with education to develop behavioral intelligence, the analysis system has a certain positive effect on helping the reform of classroom education. This study proposes an improved SSD behavior recognition model. The network model is optimized and the model convergence speed is accelerated based on the RMSProp optimization algorithm. Through a database of 2,500 images of five behaviors, including raising hands, sitting up, writing, sleeping, and playing with mobile phones, and using them as object detection datasets, we use the OpenCV library to extract frames from classroom screen recording videos as image data sources for student behavior recognition and face recognition. Finally, an improved method is proposed to change the virtual network to MobileNet and complete the fusion function. The results show that compared with the traditional SSD method, the improved model has a significantly improved effect in recognizing small objects and the recognition speed is not significantly reduced.

1. Introduction

Nowadays, legal education is mainly based on teaching in class. In the traditional teaching model, the students are passive and have scattered knowledge. The cultivation of character, emotion, ability, and other aspects has not achieved the best effect. The various disadvantages appear gradually and cannot adapt to the requirements of modern society [1]. The legal teaching in the class includes civil law, criminal law, labor law, and other law departments. The knowledge is scattered and complex, lacking the complete knowledge structure system that cannot be systematically in-depth learning, resulting in the teaching in class being still at the surface level of understanding. The learning effect has not been substantially changed [2]. With the acceleration of the popularization of higher education in China, knowing

how to guarantee and improve the quality of teaching and ensure the quality and scale of coordinated development is one of the main problems faced by Chinese colleges and universities.

Under this profound background, it is an effective way to comprehensively deepen the curriculum reform to carry out the construction of a classroom mode of promoting in-depth learning, research methods and strategies based on questions and tasks and summarize the application rules of in-depth knowledge [3]. The shallow monologue and infusing filler in class are gradually replaced by new paradigms such as cooperative inquiry and in-depth dialogue and communication. The new paradigms with exchange and communication as the central theme make classroom teaching achieve a qualitative leap. Deep learning emphasizes the deep digging of knowledge and the inquiry into the nature of things for

students, which requires the students to master the connection between knowledge generation and the knowledge system. It focuses on cultivating critical thinking and problem-solving skills for the students, which are necessary to adapt to society and still have learning abilities after entering the organization [4, 5].

As the main body of classroom teaching, their classroom behavior reflects their acceptance of knowledge and directly affects the teaching quality of teachers. The traditional education in the classroom and the status of the students can only be understood through the teacher's observation in the class, the workload is enormous, and the results are relatively one-sided. The target detection method is used to identify student behavior that can count the number of each behavior quickly in class, which is convenient and has high accuracy compared with the manual method. Therefore, the deep learning that is used to analyze classroom status and the development of artificial intelligence to improve teaching quality is the critical direction of educational research in the future, which has significant research value.

2. Related Work

The theoretical research of deep learning and classroom practice analysis is carried out almost simultaneously in foreign countries. The academic study of deep learning provides a strong foundation for classroom teaching research. Meanwhile, classroom practice research also directly enriches the theoretical investigation of deep understanding.

Rushon [6] investigated the influence of assessment methods on deep learning and highlighted the vital role of formative assessment in promoting deep understanding. The use of formative assessment for teachers to encourage students' deep learning was advocated. Hornby et al. [7] experimented on the first-year students for three cycles, and the design philosophy of Meyers and Nulty was adopted. Through the comparison between the experimental group and the control group many times, improvement and adjustment updates proved that deep learning could improve students' learning effects. Khosa et al. [8] conducted, through experimental research, pre-test and post-test that were arranged in the form of questionnaires, the promotion effect of collaborative learning on deep learning was discussed, and students were advocated to achieve the goal of deep learning through collaborative learning. Aubrey and Raath [9] clarified the learning degree through case studies that deep learning and shallow learning under problem learning mode could promote geography standard university students to achieve the goal of deep understanding. Nina [10] guided critical thinking by enhancing the interaction between teachers and students in the network environment. The research focuses on deep learning practice in the classroom. James and Richard [11] designed a comprehensive evaluation method based on constructivism to guide and promote deep learning.

Above all, deep learning mainly focuses on classroom teaching in the practical application and technical support of deep learning research in foreign research. The experimental application research had an extended period. Though the

repeated verification experiment to reach appropriate strategy and methods can promote deep learning and provide a reference for the development of classroom teaching, the support of technology for deep understanding was mainly reflected in deep knowledge in the information environment, which responded to the requirements of learning levels in the information age and promoted the effective development of deep learning for higher education learners.

Unlike foreign studies, the research of deep learning in China only stayed at the stage of logical thinking in philosophy. Deep knowledge was combined with practical teaching only in recent years.

Xu [12] combined textbook drama with deep learning. It expounds on the primary connotation of deep learning and puts forward effective strategies to promote deep learning. Wu [13] took political teaching as an example, which pointed out that teachers play an essential role in promoting deep learning and putting forward a teaching model and teaching strategy based on deep learning. Zhang et al. [14] compared deep learning with shallow learning and concluded with the critical understanding that deep learning advocated active learning, lifelong learning, and emphasis on knowledge. Zeng and Dong [15] found that deep learning was mainly "deep" in three aspects:

- The achievement of training objectives and results
- The level of thinking processing
- The level of multidimensional input

Guo [16] carried out a series of research on depth teaching and proposed that depth learning cannot be separated from the guidance and help of teachers and stressed that learners should carry out depth learning under the direction of teachers.

Through the research on the teaching practice in the classroom of deep learning in China, although existing studies have noticed the critical role of the school in promoting deep understanding, most of the research remained at the descriptive level, lacking the support of theories and experiments. Based on MobileNet's deep separable convolution structure and feature fusion theory, this paper proposed an improved SSD algorithm, constructed behavioral data sets, and trained the improved SSD network model. The SSD and enhanced models were used to identify the five behaviors of sitting in class, raising hands, writing, sleeping, and playing with mobile phones, and the recognition results were compared. The research aimed to respond to the requirements of learning levels in the information age and promote the effective development of deep learning for higher education learners.

3. The Theory and Technology of Classroom State Analysis

3.1. The Deep Learning Theory

3.1.1. Activation Function. Neural networks are commonly used in deep learning networks, the smallest of which are

neurons, in which two functions are linear and nonlinear, respectively. The former output has no relationship with the number of layers and is always linear, so the use field is limited[17]. However, in reality, neural networks are required to deal with various nonlinear problems, so a function is used to activate the results, and the web, after processing, can solve nonlinear problems. The calculation process of the activation function is shown in Figure 1, the input is put into the neuron, and the neuron performs the linear calculation on it and outputs it into the activation function; thus, a nonlinear result can be obtained [18]. The neural network with an activation function is more powerful.

(1) *Sigmoid Function*. The sigmoid function originated in the biological field, and another name for it is the logistic function. The output is controlled between 0 and 1, so it is used to activate the output results of the network layer [19]. The formula is as follows: The sigmoid function is often used in dichotomous problems, and the results are better in some aspects, but the derivation calculation is cumbersome, and the phenomenon of disappearing gradient exists.

$$f(z) = \frac{1}{1 + e^{-z}}, \quad (1)$$

where $f(z)$ is the loss function and z represents the input value.

(2) *ReLU Function*. The linear rectifying function is widely used in image recognition and computer vision. The output is the maximum value of input X and 0. The formula is as follows. Compared with the Sigmoid function, this method has complex power operation and fast calculation speed. In the calculation process, some neurons will be set to 0 to make the network become sparse and effectively reduce the overfitting phenomenon[20].

$$f(x) = \max(x, 0), \quad (2)$$

where $f(x)$ is the loss function and x represents the input value.

(3) *SoftMax Function*. This function, also known as the normalized exponential function, computes the output between 0 and 1, and the sum of the probabilities of all the outputs is 1. The formula is as follows: The SoftMax function works well for multicategory tasks but not well for keeping the same categories close together and separating different categories.

$$f(x)_j = \frac{e^{x_j}}{\sum_{k=1}^k e^{x_k}} \quad (j = 1, 2, \dots, k), \quad (3)$$

where $f(x)_j$ is the loss function, x_j is the JTH value of the input, and k is the number of input values.

3.1.2. Convolutional Neural Network. Convolutional neural networks, also known as convnets, are the foundation of deep learning, and their design is based on biological ideas. They refer to areas of interest in an image. The primary network structure is divided into three layers: input data, output data, and the middle layer. The input layer can

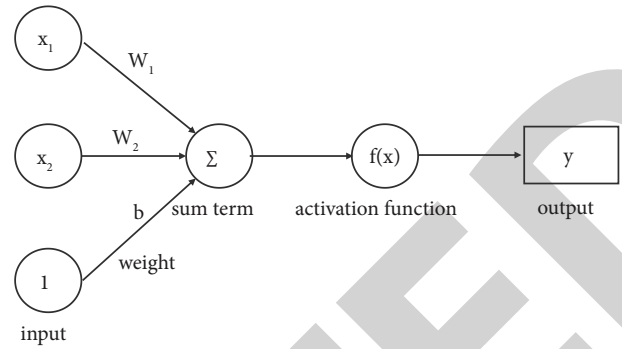


FIGURE 1: The calculation process for the activating function.

process multidimensional data before data are input to the network, which should be unified in the channel, time, and frequency. The output layer outputs corresponding results for different problems, and the middle layer is divided into the convolution layer, pooling layer, and whole connection layer.

The main mechanism of convolutional neural networks is that when the network is connected to a network, two neurons are related, but convolutional neural networks are only partially protected. I have an $N-1$ layer of neurons connected to N layers of neurons [21–25]. As shown in Figure 2, the special part connects layer $N-1$ and layer N . The left is the fully connected mode and the right is the partially connected mode. The parameter on the right is much smaller than the parameter on the left.

3.2. Face Recognition Method. MTCN face detection algorithm, affine transform face alignment, and NSightface face comparison algorithm. The face data set is first prepared, and the MTCNN algorithm and affine transformation complete face detection and alignment. Finally, the aligned faces were put into InsightFace for face comparison, and the results were obtained.

The face recognition algorithm refers to the recognition of video images in the face recognition model extracted by artificial features and the model selected in this paper. The comparison result of the artificial feature extraction method is 0.9, and the comparison result of MTCNN-InsightFace is 0.95. The accuracy of the MTNC-InsightFace face recognition model is 4% higher than that of the manual feature method, and the speed is much faster, which fully meets the needs of the actual attendance function.

3.3. Image Preprocessing Method. The influence of image noise will affect the information effect, so to ensure the required quality of the image used for operation, it is necessary to deal with it before using it. The commonly used methods include denoising, histogram equalization, normalization, and grayscale image generation.

3.3.1. Grayscale Method. The RGB model color images can get more than 16 million values for one pixel, which will increase the workload of image recognition processing.

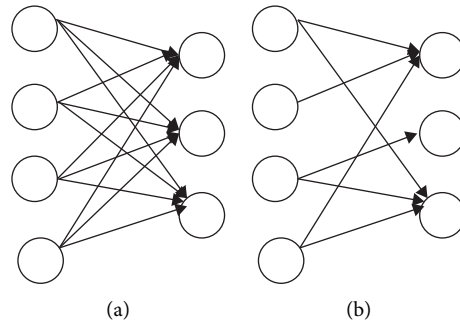


FIGURE 2: Comparison of full connection (a) and local connection (b).

Graying is $R = G = B$, ranging from 0 to 255 for a single pixel. There are many kinds of grayscale methods, mainly by seeking the average value of each pixel to grayscale, seeking the maximum value of pixels to grayscale, and seeking the weighted mean value of pixels to grayscale, the corresponding function in OpenCV tools is often used.

3.3.2. Denoising Method. The standard denoising methods include mean filter, median filter, Gaussian filter, and bilateral filter. Mean filtering belongs to linear filtering. The average of each pixel value and the surrounding pixel value is taken as the pixel value of the point. Mean filtering will destroy image details in denoising and is suitable for Gaussian noise. Median filtering is to sort the values of a pixel point and the surrounding pixels and find the intermediate value as the final pixel value, which is suitable for denoising impulse noise. The goal of Gaussian filtering is to take the mean of each pixel and the surrounding pixels after weighting and replacing the original pixel. The weight of the center point of Gaussian filtering is more significant than that of the periphery, highlighting the key points more than that of the mean filtering. However, only the spatial distance of pixel value is considered without considering the similarity, which is sufficient to cause image blur. Bilateral filtering takes spatial distance and similarity as indexes and can protect edge characteristics while removing noise.

4. Improvement of the SSD Behavior Recognition Algorithm

The behavior in class helped analyze the quality of listening and teaching effects. Five common postures, including sitting in class, raising hands, writing, sleeping, and playing with mobile phones, were selected for identification and research. The shortcomings of the target detection SSD algorithm were obtained by analyzing the characteristics of students' behavior, and an improved SSD algorithm was proposed. Meanwhile, the implementation process of the classroom behavior recognition model was introduced in detail.

4.1. The Construction Process of the Recognition Model for Behaviors in Class. The model construction mainly includes determining network structure, preparing training data,

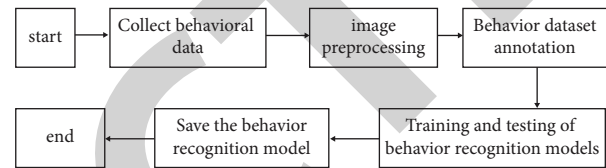


FIGURE 3: Identification process of behavior in class.

model training, and testing, etc. The behavior recognition model design process was obtained as shown in Figure 3. [2, 22, 26].

The first step was to prepare student behavior images. Two thousand five hundred images were collected, including raising hands, sitting, writing, sleeping, and playing with mobile phones, with 500 pictures for each behavior. Then, they built a behavior recognition database. The collected 2500 images were preprocessed and labeled, and the photos were divided into three parts: a training set, a test set, and a verification set according to proportion. Finally, the model was trained and tested. The initial model was obtained by putting the training set into the behavior recognition network model for training. The validation set was used to verify the model, and the network model parameters were adjusted according to the verification results. We put the test data into the model to get the results, analyze the results, and determine any differences from expectations. We decide whether to continue the training model according to the comparison results and saved the behavior recognition model with a better effect for subsequent class behavior recognition.

4.2. Improved SSD Algorithm. The improvement strategies were proposed for the primary network and small target detection of the traditional SSD algorithm. Instead of VGG16, a lightweight network was used to reduce the number of parameters to improve the detection speed. The high-level semantics were fused to the low level to enhance the small target detection effect. The improvement principle and process will be introduced in detail as follows:

4.2.1. Improvements to the Underlying Network. The goal of improving the primary network was to replace the original backbone network, VGG16, with a lightweight network. The

TABLE 1: Comparison of VGG16 and MobileNet in the imagenet dataset.

Model	ImageNet accuracy (%)	Million mult-adds	Million parameters
MobileNet	70.6	569	4.2
VGG16	71.5	15300	138

MobileNet network used deep devolution instead of ordinary convolution to reduce the number of parameters. MobileNet had only 4.2 million parameters compared with 133 million parameters in VGG16. By analyzing the test results of both in the ImageNet data set, as shown in Table 1, the speed of MobileNet was greatly improved. At the same time, the accuracy was only 0.9 percentage points lower than that of VGG16. Therefore, based on the original MobileNet, this paper was used as SSD’s primary network after some modifications.

(1) *Improvement of MobileNet.* MobileNet was faster and less computational than VGG16 because it had two differences. First, depth separable convolution was used for network composition. On the other hand, the width coefficient and resolution coefficient were also used. When the image was input into the network, a set of graphs containing feature information should be obtained through deep convolution operations, respectively, and some other feature graph information should be received by point convolution operations after the BN and ReLu operation of the feature graph, and then, the results should be obtained through BN and ReLu operations again. We change the MobileNet input size from 224×224 to 300×300 . In order to increase the information capacity of the feature map and improve detection accuracy, increasing the input size can make essential preparation for the combination of the two networks.

(2) *Replacement of SSD Primary Network.* The first 14 improved deeply separable convolutional layers were selected from the improved MobileNet(300×300) network to replace VGG16 as the backbone network of the enhanced algorithm. In order to increase the feature extraction capability of the model, eight ordinary convolutional layers of decreasing size were connected behind the replaced primary network in order to further obtain deeper information about the image. The size of the eight convolutional layers is shown in Table 2.

Finally, a classification layer for judging categories and a nonmaximum suppression layer for screening regression boxes were connected to the end of the network to complete the replacement of the primary network. The basic network structure after the replacement is shown in Table 3. Deep convolution and subsequent 1×1 point convolution were regarded as one layer, and there were 14 layers, respectively, denoted as Conv0 to Conv13, where s_1 represents step size 1, s_2 represents step size 2, and Conv DW represents deep convolution and then convolved with a 1×1 point to process the channel.

The same as with the original SSD, 6 feature layers were selected to complete feature extraction and target detection. The depth of layers was considered in the selection, as it was

too shallow to extract enough image information. The 6 characteristic layers selected in this paper were Conv11, Conv13, Conv14_2, Conv15_2, Conv16_2, and Conv17_2, which decrease in size from front to back to achieve multi-scale prediction.

4.2.2. *Feature Fusion of the Network Model.* The replacement of a primary network improved the detection speed but did not improve the accuracy of small target detection. Improving model performance by integrating features of different scales was the common improvement strategy [9, 27].

Combined with the structure of the network model and the characteristics of each feature fusion method, the ADD feature fusion method was selected by the researchers for the network fusion operation. During the fusion, the fusion layer was selected first. Conv17_2 and Conv16_2 were too small to have much information. Only Conv11, Conv13, Conv14_2, and Conv15_2 were chosen for the fusion operation. The specific fusion steps were as follows:

In the first step, the size of Conv15-2 changed from 3×3 to 5×5 after up-sampling, and then, it was fused with Conv4-2 by the add method. Finally, the fusion results were normalized to obtain the characteristic layer.

The second step: an upsampling operation was carried out on conv14_2_r with the size of 5×5 to make it the same size as Conv13 with the size of 10×10 . The two were fused and normalized by the ADD method to obtain the new feature layer CONV13_r.

The third step: the feature layer conv13-R obtained in the previous step was up-sampled to receive the exact size of 19×19 as Conv11. The add feature fusion method was also used for fusion, and finally, normalized.

4.2.3. *Model Optimization Algorithm.* Model training requires continuous attention to the change of loss function. The constant decrease in loss function value indicates that the result of model training was closer to the natural consequence. In order to accelerate the decline speed, optimization algorithms were usually used, such as Momentum, RMSprop, and Adam. This paper adopts the RMSProp (Root Mean Square Prop) optimization algorithm proposed by Geoffrey E. Hinton. In this algorithm, the historical gradient of each dimension was squared and superimposed. The decay rate was introduced simultaneously to obtain the sum of the historical rise. The learning rate was divided by the result obtained above when the parameters were updated. After using the algorithm, the gradient direction changes in a small range, which speeds up the convergence of the network. The specific calculation formula was shown as follows:

TABLE 2: Size of eight convolution layers.

Name of the layer	Conv 14-1	Conv 14-2	Conv 15	Conv 15-2	Conv 16-1	Conv 16-2	Conv 16-1	Conv 17-2
Layer size	10×10	5×5	5×5	3×3	3×3	2×2	2×2	1×1

TABLE 3: Basic network structure after replacement.

Convolutional layer name	Convolution method/step length	Convolution kernel shape	Input size
Conv0	Conv/s2	$3 \times 3 \times 3 \times 32$	$300 \times 300 \times 3$
Conv1	Conv dw/s1	$3 \times 3 \times 32dw$	$150 \times 150 \times 32$
	Conv/s1	$1 \times 1 \times 32 \times 64$	$150 \times 150 \times 32$
Conv2	Conv dw/s2	$3 \times 3 \times 64dw$	$150 \times 150 \times 64$
	Conv/s1	$1 \times 1 \times 64 \times 128$	$75 \times 75 \times 64$
Conv3	Conv dw/s1	$3 \times 3 \times 128dw$	$75 \times 75 \times 128$
	Conv/s1	$1 \times 1 \times 128 \times 128$	$75 \times 75 \times 128$
Conv4	Conv dw/s2	$3 \times 3 \times 128dw$	$75 \times 75 \times 128$
	Conv/s1	$1 \times 1 \times 128 \times 256$	$38 \times 38 \times 128$
Conv5	Conv dw/s1	$3 \times 3 \times 256dw$	$38 \times 38 \times 256$
	Conv/s1	$1 \times 1 \times 256 \times 256$	$38 \times 38 \times 256$
Conv6	Conv dw/s2	$3 \times 3 \times 256dw$	$38 \times 38 \times 256$
	Conv/s1	$1 \times 1 \times 256 \times 512$	$19 \times 19 \times 256$
Conv7	Conv dw/s1	$3 \times 3 \times 512dw$	$19 \times 19 \times 512$
	Conv/s1	$1 \times 1 \times 512 \times 512$	$19 \times 19 \times 512$
Conv8	Conv dw/s2	$3 \times 3 \times 512dw$	$19 \times 19 \times 512$
	Conv/s1	$1 \times 1 \times 512 \times 512$	$19 \times 19 \times 512$
Conv9	Conv dw/s1	$3 \times 3 \times 512dw$	$19 \times 19 \times 512$
	Conv/s1	$1 \times 1 \times 512 \times 512$	$19 \times 19 \times 512$
Conv10	Conv dw/s2	$3 \times 3 \times 512dw$	$19 \times 19 \times 512$
	Conv/s1	$1 \times 1 \times 512 \times 512$	$19 \times 19 \times 512$
Conv11	Conv dw/s1	$3 \times 3 \times 512dw$	$19 \times 19 \times 512$
	Conv/s1	$1 \times 1 \times 512 \times 512$	$19 \times 19 \times 512$
Conv12	Conv dw/s2	$3 \times 3 \times 512dw$	$19 \times 19 \times 512$
	Conv/s1	$1 \times 1 \times 512 \times 1024$	$10 \times 10 \times 512$
Conv13	Conv dw/s1	$3 \times 3 \times 1024dw$	$10 \times 10 \times 1024$
	Conv/s1	$1 \times 1 \times 1024 \times 1024$	$10 \times 10 \times 1024$

$$S_{dR} = \beta S_{dR} + (1 - \beta)(dR)^2, \quad (4)$$

$$R = R - \rho \frac{dR}{\sqrt{S_{dR} + a}}$$

where, β is the rate of decay, S_{dR} is Cumulative gradient variable, ρ is learning rate, a is the constant that is not zero, and R is the parameter.

4.3. Behavioral Database Building Methods. The good classification effect of the deep learning network model should be based on a large number of data, which constitute an image database. At present, there was no database specially used for classroom behavior recognition.

4.3.1. Data Set Acquisition and Enhancement. The data set came from classroom surveillance videos and network pictures. In order to ensure the recognition effect, the video images need to be processed before being used as the data set, from which video segments including raising hands, sitting up, sleeping, and writing were selected. OpenCV was used for frame sampling of the selected video, and the

pictures containing the above five actions were selected for saving. 1000 image data were collected in this experiment, and some graphic data are shown in Figure 4.

The precision of model training needed a large amount of data as support, so data enhancement was used to increase the amount of data, which included flipping the image horizontally, left-right, and randomly, translating the image horizontally and vertically, and randomly changing the color of the image. After data enhancement, the dataset for this paper contains 2500 images.

4.3.2. Data Set Preprocessing. The collected color image would increase the model trained workload, and the image was prone to contain noise due to the influence of the external environment, so the data set needs to be processed by grayscale and denoising methods, and sharpened by the object enhancement method.

(1) Grayscale processing. The mean value of each pixel point was calculated to realize grayscale processing. The calculation formula was as follows. The comparison before and after gray processing used the formula as shown in Figure 5.



FIGURE 4: Comparison of NMS effects on graphic data (hand up, mobile phone, sleep, lecture, writing).

$$R = G = B = \frac{(R + B + G)}{3}, \quad (5)$$

where, R , G , and B are three color channels.

(2) *Bilateral filtering denoising technology*. The bilateral filtering denoising technology was adopted. In an operation similar to Gaussian filtering, each pixel of the image was scanned once, and the weighted sum of the pixel values and corresponding position weights was added on the basis of the operation of obtaining the weighted sum of each pixel value in the field and corresponding position weights. In the calculation, the closer the center was, the greater the weight was, and the closer the pixel value was, the greater the weight was. The specific formula was as follows.

$$G_s = \exp\left(-\frac{\|p - q\|^2}{2\sigma_s^2}\right), \quad (6)$$

$$G_r = \exp\left(-\frac{\|I_p - I_q\|^2}{2\sigma_r^2}\right),$$

where G_s is the spatial distance weight, G_r is the Pixel weight, q is the central store of Window, p is any point, I_q is the input image, and I_p is the filtered image.

$$W_q = \sum_{p \in S} G_s(p) G_r(p)$$

$$= \sum_{p \in S} \exp\left(-\frac{\|p - q\|^2}{2\sigma_s^2}\right) \exp\left(-\frac{\|I_p - I_q\|^2}{2\sigma_r^2}\right), \quad (7)$$

where W_q is the sum of the weights of each pixel value.

Gaussian filtering mainly played the role of image smoothing. In the critical part of the image, there would be obvious color or light and shade transformation, which was reflected at the pixel level, that was, the pixel values on both sides differed greatly, and the difference gradually increased with the distance. In this case, the G_r value was close to 0, and the whole filter result was also 0. The two images that were, respectively, processed by the method used in this paper (left) and Gaussian filtering (right) are shown in Figure 6.



FIGURE 5: Image grayscale noise comparison.



FIGURE 6: Image denoising.

(3) *Objective to enhance*. Unsharpen Mask (USM) was used to enhance the target. The input image was processed with a low-pass filter to obtain the low-pass component, and the difference between the original image and the component was calculated to obtain the high-pass component, and the sharpened image was obtained by superposition the high-pass component on the basis of the original image. The Gaussian fuzzy method was usually used to obtain low-pass components. The calculation formula was as follows.

$$y = \frac{(x - w \times z)}{(1 - w)}, \quad (8)$$

where y is the output image, x is Gaussian Blur, ranging from 0.1 to 0.9, usually 0.6, z is the weight value.

We input each pixel in the image for USM operation and obtain the pixel value after sharpening each pixel, thus forming the whole sharpened image to complete the target enhancement. The classroom behavior imaged after target enhancement is shown in Figure 7.

4.3.3. *Data Set Annotation*. Annotation tool uses the LabelImg image annotation tool. The software processes the image according to the format of the Pascal VOC data set. Before labeling, the preprocessed image needed to be saved in Pascal VOC format. After labeling, some basic information about the image, including storage location, size, and category name, was automatically saved in XML files.



FIGURE 7: Image sharpening.

TABLE 4: Experimental environment.

Operating system	Windows
Hardware environment	Intel core i5; NVIDIA GeForce GTX1080
Programming environment	Python3.5
Software engineering	Keras deep learning framework, anaconda carrying platform; OpenCV computer vision library

TABLE 5: Comparison of recognition effects of different models.

Monitoring framework	Based on the network	Whether feature fusion	Average accuracy (mAP) (%)	Detection speed (fps)
SSD	VGG16	No	84.00	22.0
MobileNet-SSD	MobileNet	No	76.14	27.1
Feature fusion of MobileNet-SSD	MobileNet	Yes	83.08	24.6

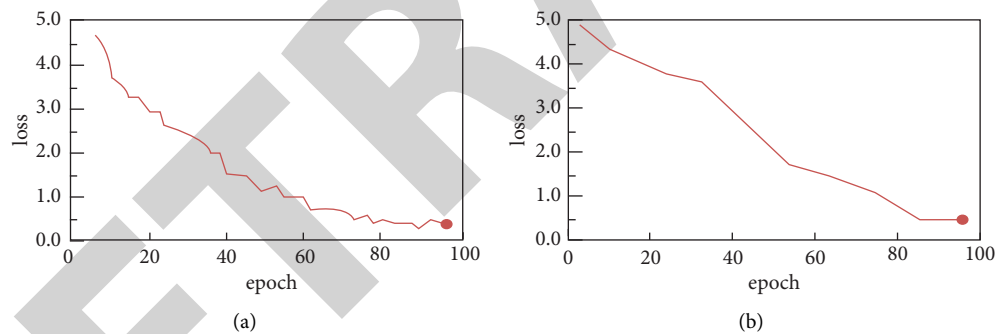


FIGURE 8: Loss variation.

TABLE 6: Comparison of the accuracy of five classroom behavior tests.

Model checking	Listen to lecture (%)	Play phone (%)	Raise hand (%)	Writing (%)	Sleep (%)
SSD	89.53	79.74	86.27	77.09	87.66
Improved model	89.31	80.15	87.76	82.12	86.04

5. Experimental Study on the Improved SSD Algorithm

The traditional SSD algorithm, the unimproved MobileNet-SSD algorithm, and the improved MobileNet-SSD algorithm were compared and analyzed from three aspects of training difficulty, detection accuracy, and detection speed.

5.1. Experimental Environment and Parameter Setting. The 500 images were selected for each action in the test training set and 2500 images were selected for each action in

the test set. During the training, batch size was set to 4, 625 batches were needed for 2500 training sets, and the epoch was set to 100, that was, 62500 iterations in total. Among them, the learning rate of the first 5000 times was 10^{-4} and the learning rate of the latter was 10^{-5} . The test environment is shown in Table 4.

5.2. Model Evaluation Criteria. In this paper, the model was evaluated by single frame image detection time and Mean Average Precision (mAP) of image detection, which was the mean value of all AP values. AP was the area below the curve

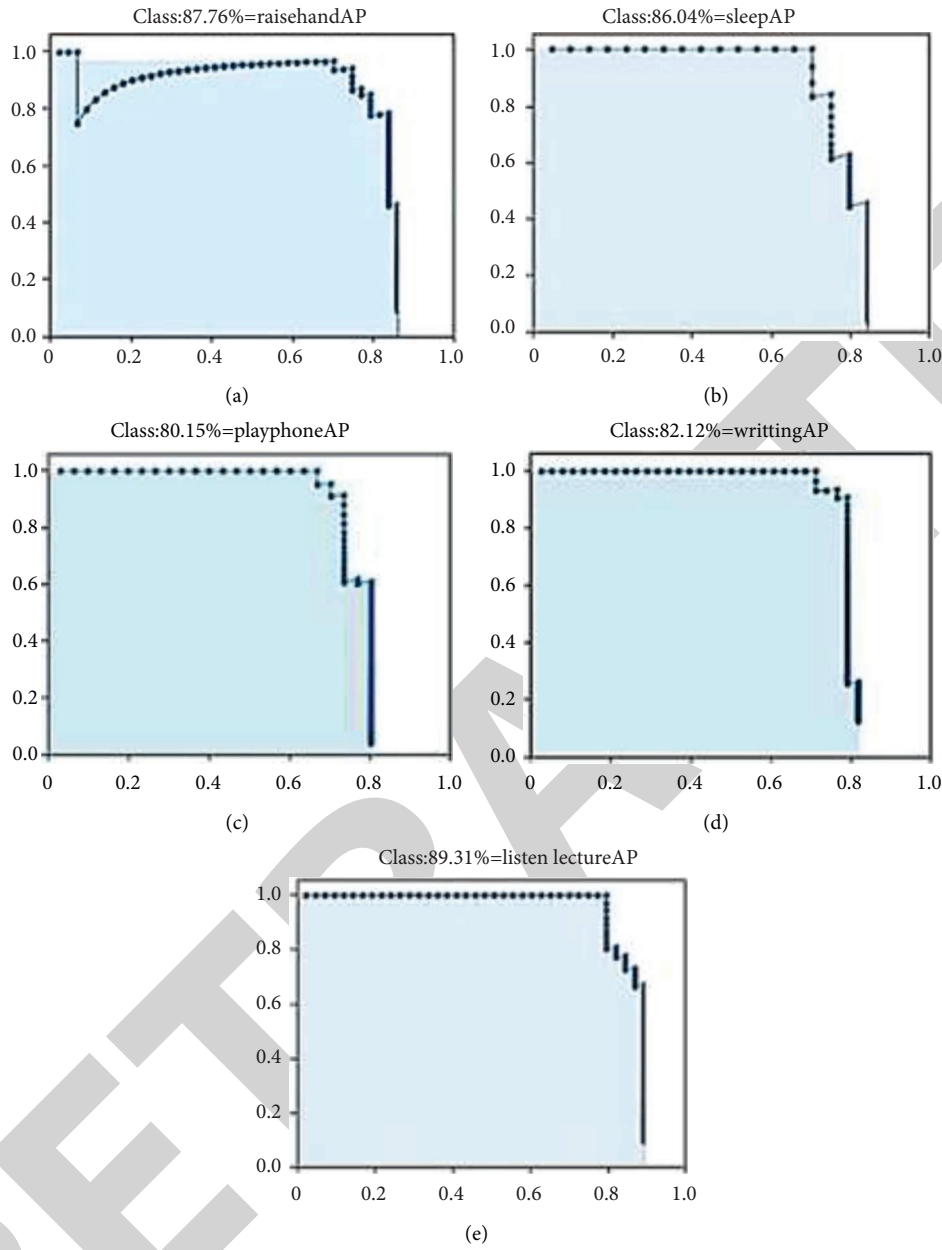


FIGURE 9: Identify accuracy.

composed of precision and recall. The formula for accuracy was as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (9)$$

where TP is the classifier that divides the target into positive samples and the number of samples that are actually positive samples and FP is the number of samples that the classifier considers positive but is actually negative.

The whole formula represents the proportion of positive samples considered by the classifier to positive samples recognized by the whole classifier, reflecting the model's precision function.

The recall rate formula was as follows:

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (10)$$

where FN is the number of samples that the classifier treats as negative samples but is actually positive samples.

The whole formula represents the proportion of the samples considered positive and confirmed positive by the classifier to all positive classes, reflecting the model's recall function.

5.3. Model Experiment Process

5.3.1. Preparation of Documents Required for the Test. Before model training, the train. txt file was generated by running the code, which was used to store the training set information.

5.3.2. The Model Training. The training started with an image file load call `annotation_path = "train.txt"` to reference the resulting train. Then, we identify the category setting. The test needed to identify five behaviors and backgrounds, a total of six categories, namely, `NUM_CLASSES = 6`. Finally, the network structure was loaded and trained, The model = `r_mSSD300()` method was used to load the improved model and then accorded to the parameter settings for training. Repetitive training can obtain a better model and save as the behavior recognition model in this paper.

5.4. The Analysis of Model Experiment Results. The traditional SSD algorithm, MobileNet-SSD, and mobile net-SSD with feature fusion were trained in the same experimental environment and parameters, and the three algorithms were compared through the test set. The data used in the test were self-made data sets. The average accuracy and detection speed (detection time per frame) of classroom behavior detection obtained by different models are shown in Table 5.

As seen from Table 4, classroom behavior recognition experiment, compared with the traditional SSD algorithm, the feature fusion MobileNet-SSD improved the detection speed by 2 frames per second, and the average detection accuracy reached 83.08%. Compared with a mobile net-SSD model without feature fusion, the speed was reduced by 2.5% and the accuracy was improved by 6.94% due to the increase in network parameters by fusion. The analysis results showed that the detection speed and recognition accuracy of the proposed algorithm was improved obviously.

The difficulty of model training can be judged by comparing the curve of the loss function during training. With the same parameters, the loss function curve of mobile net-SSD and SSD models with feature fusion was taken when the epoch was 100 iterations for 50000 times, as shown in Figure 8. The loss values of both models continued to decrease, proving that both models were reasonable. During the training, it took 6 days for the loss of the model used in this paper to drop below 0.5, while the original SSD model took 8 days. In addition, the loss value of the model in this paper decreased rapidly, so the training difficulty of the model in this paper was less than that of the traditional SSD model.

The mobile Net-SSD model with SSD and feature fusion was used to test the five actions of students in the test set: listening in class, raising hands, writing, sleeping, and playing mobile phones. The detection accuracy (AP) of each action is shown in Table 6.

It can be seen from Table 6 that compared with the original SSD algorithm, the mobile net-SSD algorithm based on feature fusion in this paper had improved the detection effect of small targets in the five actions, among which the writing improvement reached the highest level of 3.03%, indicating that the model in this paper had improved in small target recognition. Observing the mobile net-SSD recognition result of feature fusion, among the five

movements, listening accuracy was the highest, followed by raising hands, and writing and playing on mobile phones were the lowest. Through analysis, the reasons for this result were as follows: compared with other movements, the two activities were more prone to occlusion, especially in the recognition process. It was easy to be confused with other hand movements, so the correction effect was not as good as the other three movements. To more intuitively display the recognition accuracy rate of the model in this paper on the five actions, the line graph is shown in Figure 9, where the shaded area was the accuracy rate.

6. Conclusion

Because of the subjective one-sidedness of traditional legal classes in which students' state was observed artificially, intelligent analysis was introduced into the category, and the deep learning method was proposed to identify students' classroom behavior. The model algorithm was integrated into the system to expand it, and the legal class state analysis system was designed and implemented. The following conclusions were drawn:[28].

- (1) The design process of the class behavior recognition model was analyzed, and the images of listening in class, raising hands, writing, sleeping, and playing with mobile phones were selected as the student behavior database after grayscale, noise reduction, and image enhancement practice improved behavior recognition models.
- (2) Based on the principle of deeply separable convolution, the improved SSD method was used to analyze student states. Changing the original SSD base network from VGG16 to the improved MobileNet network, using add feature fusion method to replace the network, reducing the basic network parameters, and integrating in-depth information into the shallow layer, the detection effect and speed of small targets were improved.
- (3) The improved model was trained and used for student behavior recognition. We identify the population distribution of five behaviors in a class, obtain the population's proportion in the set severe, good, average, poor, and other five states, and complete the analysis test of student status.
- (4) The university class state analysis system based on deep learning helps users intuitively analyze the state information of students in a single or multiple classes which is convenient for understanding the class situation of law students and provides a reference for course adjustment.

Data Availability

The dataset can be accessed upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] M. Wang, B. Liu, and H. Foroosh, "Factorized convolutional neural networks," 2016.
- [2] N. Kashiwagi, H. Tanaka, Y. Yamashita et al., "Applicability of deep learning-based reconstruction trained by brain and knee 3T MRI to lumbar 1.5T MRI," *Acta Radiologica Open*, vol. 10, no. 6, Article ID 205846012110239, 2021.
- [3] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [4] S. Miao, Z. J. Wang, and R. Liao, "A CNN regression approach for real-time 2D/3D registration," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1352–1363, 2016.
- [5] D. Maturana and S. Scherer, "A 3d convolutional neural network for real-time object recognition," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2–11, Hamburg, Germany, September 2015.
- [6] A. Rushton, "Formative assessment: a key to deep learning," *Medical Teacher*, vol. 27, no. 6, pp. 509–513, 2005.
- [7] G. Hornby, G. Jennings, and D. Nulty, "Facilitating deep learning in an information systems course through application of curriculum design principles," *Journal of Teaching in Travel & Tourism*, vol. 9, no. 1-2, pp. 124–141, 2009.
- [8] D. K. Khosa, S. E. Volet, and J. R. Bolton, "An instructional intervention to encourage effective deep collaborative learning in undergraduate veterinary students," *Educational Methods*, vol. 37, no. 4, p. 369, 2010.
- [9] G. Aubre and S. Raath, "Problem-based learning to foster deep learning in preservice geography teacher education," *Journal of Geography*, pp. 56–58, 2015.
- [10] A. R. Nina, "Strategies for effective faculty involvement in online activities aimed at promoting critical thinking and deep learning," *Education and Information Technologies*, vol. 21, no. 4, pp. 881–896, 2016.
- [11] A. B. James and G. F. Richard, "Beyond measurement-driven instruction: achieving deep learning based on constructivist learning theory. Integrated assessment, and a flipped classroom Approach," *Journal on Excellence in College Teaching*, vol. 28, no. 2, pp. 111–132, 2017.
- [12] Z. Y. Xu, *Deep Learning in the Textbook Play*, Beijing Normal University Publishing House, China, 2012.
- [13] B. Wu, *Apply the Model Based on Deep Learning to the Political Classroom*, People's Education Press, Beijing China, 2012.
- [14] H. Zhang, X. J. Wu, and J. Wang, "Objective and evaluation system construction of deep learning," *Education and Information Technologies*, no. 7, pp. 51–55, 2014.
- [15] J. Y. Zeng and Z. H. Dong, "Research on the connotation and cultivation path of students'," *Deep Learning*, vol. 25, no. 8, pp. 59–67, 2017.
- [16] Y. Y. Guo, "Deep learning: essence and philosophy," *Survey of Education*, vol. 14, no. 1, p. 11, 2017.
- [17] S. Gassenmaier, S. Afat, D. Nickel, M. Mostapha, J. Herrmann, and A. E. Othman, "Deep learning-accelerated T2-weighted imaging of the prostate: reduction of acquisition time and improvement of image quality," *European Journal of Radiology*, vol. 137, Article ID 109600, 2021.
- [18] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu et al., "Convolutional neural networks for medical image analysis: full training or fine tuning[?]," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [19] M. Kallenberg, K. Petersen, M. Nielsen et al., "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1322–1331, 2016.
- [20] J. Masci, D. Boscaini, M. Bronstein, and P. Vandergheynst, "Vandergheynst. Geodesic convolutional neural networks on Riemannian manifolds," *Proceedings of the IEEE International Conference on Computer Vision Workshops*, vol. 2, pp. 37–45, 2015.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2015.
- [22] H. T. Nguyen, Z. K. Shah, A. Mortazavi et al., "Periodically rotated overlapping parallel lines with enhanced reconstruction acquisition to improve motion-induced artifacts in bladder cancer imaging: initial findings," *Medicine (Baltimore)*, vol. 98, no. 42, Article ID e17075, 2019.
- [23] A. A. A. Setio, F. Ciompi, G. Litjens et al., "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.
- [24] A. Eklund, P. Dufort, D. Forsberg, and S. M. LaConte, "Medical image processing on the GPU - p," *Medical Image Analysis*, vol. 17, no. 8, pp. 1073–1094, 2013.
- [25] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Isgum, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1252–1261, 2016.
- [26] T. Brosch, L. Y. W. Tang, Y. Yoo, D. K. B. Li, A. Trabulsee, and R. Tam, "Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1229–1239, 2016.
- [27] S. Gassenmaier, S. Afat, D. Nickel, M. Mostapha, J. Herrmann, and A. E. Othman, "Deep learning-accelerated T2-weighted imaging of the prostate: reduction of acquisition time and improvement of image quality," *European Journal of Radiology*, vol. 137, Article ID 109600, 2021.
- [28] D. J. Winkel, T. J. Heye, M. R. Benz et al., "Compressed sensing radial sampling MRI of prostate perfusion: utility for detection of prostate," *Cancer*, vol. 290, no. 3, pp. 337–345, 2019.