Application of SSD core detection algorithm in intelligent visual monitoring of examination room

Tao Xu\(^1\), Han Xiao\(^1\)

\(^1\)Academic affaires office, Guangzhou College of Commerce, Guangzhou 511363, Guangdong, China

*Corresponding author e-mail: xutao@gcc.edu.cn

Abstract. In the examination room monitoring, the traditional mode relies heavily on human resources, and only uses the surveillance camera as an auxiliary tool. In order to solve the problem of intelligent visual monitoring in the examination room, this paper proposes to use SSD core detection algorithm to realize real-time invigilation based on Deep-learning, and verifies the detection performance of the algorithm through data set test. The results show that the average accuracy of SSD300 algorithm is 79.8\%, the detection frame rate is 46, and the algorithm performance is the best; at the same time, the accuracy and recall rate of SSD300 algorithm are improved by 3.0\% and 3.4\% respectively. This study has achieved good results, which shows that the proposed detection algorithm and optimization ideas have a certain application value.

Keywords: Ssd Algorithm, Examination Room Monitoring, Intelligent Vision, Neural Network

1. Introduction

With the development of big data technology and the popularization of intelligent application, the intelligent ecosystem in various fields is gradually improving, and the potential of intelligent era is highlighted. The traditional invigilator mode requires one or two invigilators in each examination room, and invigilators outside the examination room, which means a lot of human resources are needed to cooperate with the invigilator work [1]. The age of intelligence provides the possibility of "human burden reduction". Machine vision technology can extract and identify the examinee's behavior characteristics, and judge the type of cheating behavior by boundary classification. But the traditional target detection algorithm also has some problems, such as the sliding window takes up a large amount of memory, takes a long time and so on [2]. The development time of Deep-learning is short, but its design idea is simple and effective, and it leaps and develops with the support of big data technology, which has become one of the important directions of the development of the future intelligent era [3]. This research will explore the Deep-learning technology based on regression and its application effect in real-time target detection, aiming to provide reference for intelligent visual monitoring of examination room.
2. **Research on cheating detection in examination room based on SSD core detection algorithm**

Deep-learning algorithm is an end-to-end network model based on a large number of data. In essence, it is a multi-layer neural network, which abstracts and expresses knowledge through fully automatic learning, so as to realize the high concentration and generalization of the original data [4]. Region Proposal replaces the sliding window in traditional target detection, and can get higher quality target features. Although this method improves the processing accuracy of subsequent convolutional neural network, its efficiency is difficult to meet the real-time requirements. Relatively speaking, using regression method to extract the features of the target object from the image or video frame can achieve the effect of real-time detection. YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector) are two representative regression detection methods. They are all one-stage methods, that is, selecting different scales and aspect ratios, uniformly and densely sampling on multi-layer image features, and then using convolutional neural network for feature classification and regression. Its advantage is fast running speed, but even dense sampling means that the positive and negative samples are not balanced, so it is difficult to train the neural network, which will affect the accuracy of the algorithm [5].

![Figure 1. Schematic diagram of YOLO network structure](image)

Figure 1 is the schematic diagram of YOLO network structure. Firstly, the input image is divided into 7*7 grids; Secondly, the possibility of each grid containing the target and its category is predicted twice; the predicted target window is filtered to remove the low confidence window; Finally, the redundant window is eliminated by non maximum suppression. This direct regression method can directly obtain the target location and category, and the process is simpler. According to the model structure in Figure 1, it contains 24 convolution layers and 2 fully connected layers, and there are two sizes of kernel functions in convolution layer. YOLO network improves the last two full connection layers. It adds a 7*7*30 dimensional convolution layer after the 4096 dimensional full connection layer, which makes the full graph feature regress the feature information of each grid again. On the other hand, the disadvantages of YOLO network are obvious. Firstly, the prediction part of the algorithm is only executed twice, and the two predictions are of the same kind, so the features with high similarity can not be distinguished; secondly, the generalization ability of YOLO algorithm is weak; Thirdly, the positioning error caused by its loss function will affect the accuracy of the algorithm.
Figure 2. network architecture of SSD300

SSD algorithm discretizes the picture bounding box to get a set of default boxes. These default boxes with different sizes may contain different potential targets. The algorithm will predict them separately and adapt to the target shape dynamically. SSD algorithm is consistent with the regression idea of YOLO algorithm, and it also draws lessons from the Anchor mechanism of Faster-RCNN [6]. According to the size of the input image, SSD algorithm can be divided into SSD300 and SSD512. Figure 2 shows the network architecture of SSD300. The convolution layer of feature extraction uses VGG16 structure for reference. SSD algorithm has three core concepts: one is to use multi-scale feature map, the other is to set Default boxes, and the third is to use convolution to extract target features. After determining the training samples, we need to select the loss function, which is defined as the weighting of the confidence error and the position error, as shown in formula (1).

\[ L(x,c,l,g) = \frac{1}{N} (L_{conf}(x,c) + L_{loc}(x,l,g)) \]  

(1)

In equation (1), \( N \) is the number of positive samples in the prior box, \( c \) is the confidence of the prediction category, \( l \) is the predicted position value of the corresponding boundary of the prior box, and \( g \) is the position parameter of the ground truth box. After cross validation, the weight coefficient \( \alpha = 1 \) is set.

In addition, in order to improve the efficiency of SSD algorithm for target detection, Triple Loss based on metric learning is used to optimize the algorithm. Its essential purpose is to minimize the feature distance between positive samples and class samples, and maximize the feature distance between positive samples and negative samples [7]. In this paper, the interval parameter \( \alpha \) of the distance between the same class and the distance between different classes is proposed. The optimization function is shown in equation (2).

\[ \min_w \frac{1}{m} \sum_{a} \| f(x^a) - f(x^p) \|^2_2 + \alpha \]  

(2)

In equation (2), \( a \) is positive sample, \( p \) is class sample, \( n \) is negative sample, and \( \| \cdot \|_2 \) is Euclidean distance measure. Considering the concealment of cheating behavior, the algorithm is equivalent to comparing the similar images in target detection, so let \( \text{margin} = 0.5 \).

Different from other behavior detection, the cheating behavior of examinees is very hidden, which means the need for intelligent recognition of images with high similarity. On the other hand, although there are many public behavior data sets for intelligent monitoring, there is no special data set for cheating recognition training in the examination room, so this study needs to build a special dataset.

In the case of excluding substitute test and fake qualification, we only consider monitoring the cheating behavior of examinees. The cheating behavior of examinees is divided into three categories, namely, using mobile phone/copy, peaking at the answers, and passing the answers. Before the experiment, the cheating invigilator videos in different scenes and angles are collected; in order to ensure that the key feature information is not lost, the video needs to have high enough resolution. The video is interpolated in a single frame, and one image is saved every 7 frames. A total of 10927 original images are obtained. The image pixels are 1280*720 and saved in 24 bit JPEG format. Before algorithm training, the image is labeled manually to distinguish normal behavior from cheating behavior. Among them, the normal behaviors include students' behaviors other than cheating and teachers' walking behaviors. In order to ensure the training effect, the labeled original image information is enhanced, and a total of 40000 behavior samples are obtained, in which the normal behavior: cheating behavior was 3:1. At the same time, all the behavior samples were classified manually in advance to verify the correctness of the follow-up experiments. The classification tags were normal behavior, peeping at mobile phones, peeping at others and passing answers.

3. SSD core detection algorithm application effect verification
3.1. Target detection performance verification of different algorithms
Mean Average Precision (MAP) is selected as the evaluation index of algorithm precision and the
detection speed of the algorithm is evaluated by Frame Per Second (FPS). According to the
classification, the constructed behavior sample set is divided into training set and test set in a ratio of
3:1. The training set is used to train Faster R-CNN algorithm, YOLO algorithm, SSD300 algorithm
and SSD500 algorithm, and the test set is used to test the detection effect of the algorithm, as shown in
Figure 3.

![Figure 3. detection index distribution of different algorithms](image)

Figure 3 shows the detection index distribution of different algorithms. From the frame rate FPS
point of view, the speed of YOLO algorithm and SSD300 algorithm are at a high level, in which the
FPS value of SSD300 is 46; the FPS value of SSD512 algorithm is in the second gradient, and the last
is the Faster R-CNN algorithm; from the MAP value point of view, the accuracy level of the two SSD
algorithms is higher, above 77%; the accuracy level of YOLO algorithm is the lowest, only 63.4%.

3.2. Application of SSD core detection algorithm in examination room scene
In order to verify the real-time application effect of the algorithm in the actual invigilator scene, we
choose the YOLO algorithm and SSD300 algorithm to test. The experiment is carried out under the
Caffe framework, and the test results are shown in Figure 4. From the detection results in Figure 4,
SSD algorithm can effectively detect cheating in three scenarios; although YOLO algorithm also
detects cheating in scenarios 1 and 2, it fails to detect cheating in scenario 3. This test result is also
consistent with the precision performance test result in Figure 3. From this point of view, although
both YOLO algorithm and SSD algorithm can be used for real-time detection, YOLO algorithm may
have missed detection.
Based on the traditional SSD300 algorithm, this study uses formula (2) to optimize. The detection results of SSD algorithm before and after optimization are shown in Table 1.

**Table 1.** Test results of SSD algorithm before and after optimization

| SSD300 algorithm     | accuracy/% | recall/% |
|----------------------|------------|----------|
| Before optimization  | 81.3       | 76.8     |
| After optimization   | 84.3       | 80.2     |

From the results in Table 1, the accuracy and recall of the optimized algorithm are improved by 3.0% and 3.4% respectively. From this point of view, the optimization ideas proposed in this study have good effect and certain application value.

**4. Conclusion**

The intelligent development of the whole national education system is inseparable from the intelligent transformation of each small system, which not only responds to the development trend of the times, but also greatly reduces the labor cost and improves the objectivity of the system operation. Based on the Deep-learning framework, aiming at the problem of intelligent visual monitoring of examination room, this paper proposes a cheating detection algorithm with SSD algorithm as the core, and optimizes it; through the comparison with other algorithms, the application effect is verified. From the research results, the FPS value and MAP value of SSD300 are the best, which are 46 and 79.8% respectively; compared with the YOLO algorithm, the detection effect is better; at the same time, the accuracy and recall rate of the optimized SSD300 algorithm are improved by 3.0% and 3.4% respectively. This research has achieved good results, which also means that Deep-learning has a
certain application prospect in real-time invigilation. However, looking back on the whole research process, there are many shortcomings, Such as the research is only verified by data sets, but it is not applied in actual scenarios. In addition, the detection algorithm proposed in this study has more room for improvement, which not only depends on the development of machine learning technology, but also needs to deeply integrate the thinking of Deep-learning into the solution of practical problems.

References
[1] Su Juan, Yang long, Huang Hua, et al. Improved SSD algorithm for small target ship detection in SAR images [J]. Systems engineering and electronics, 2020, 42 (5): 1026-1034.
[2] Wu Huayun, Ren Dejun, Fu Lei, et al. Surface defect detection of empty bottles based on improved SSD algorithm [J]. Computer and modernization, 2020, No. 296 (4): 125-130.
[3] Lei Huadi, Chen Dongfang, Wang Xiaofeng. Target detection algorithm based on cascaded SSD [J]. Computer engineering and design, 2020, v.41; No.408 (12): 225-232.
[4] Zhang Zengguang. Research on moving object detection method in intelligent video surveillance system [J]. Modern scientific instruments, 2019 (4): 67-70.
[5] Han Jialin, Wang Qiqi, Yang Guowei, et al. SSD network compression combining weight and convolution kernel pruning [J]. Computer science, 2019, 46 (11): 272-276.
[6] Chen Qianjun. Ship false target image detection algorithm in intelligent video surveillance [J]. Ship science and technology, 2019, 41 (2): 206-208.
[7] Chen Huan Jie, Wang Qi Qi, Yang Guo Wei, et al. SSD target detection algorithm based on multiscale convolution feature fusion [J]. Computer science and exploration, 2019,13 (6): 1049-1061.