Detection of Prohibited Articles based on Lightweight Convolutional Neural Network

Xintao Lin¹, a, Yanxi Zhang¹, b, Nanfeng Zhang², c, Xiangdong Gao¹, *, Jieshan Ruan², d

¹Guangdong Provincial Welding Engineering Technology, Research Center, Guangdong University of Technology, Guangzhou, China
²Huangpu Customs Technical Center, Dongguan, China

¹1312707786@qq.com; b15215564@qq.com, c253642845@qq.com; d2296381326@qq.com;
* Corresponding author: gaoxd666@126.com

Abstract—Prohibited articles detection is an important way to ensure public safety. In order to improve the flexibility of security inspections in complex public places, this paper uses three different convolutional neural networks, YOLOv3, MobileNet-SSD, and YOLOv3-Tiny, to detect the prohibited articles on the security inspection image data. Two performance indicators, mean Average Precision and Frames Per Second, were used to test the effect of the propose model. Finally, this trained model was deployed to the Raspberry Pi mobile device. Experimental results show that using YOLOv3-Tiny model for prohibited articles detection can improve the detection speed by about 10 times and the high detection precision is also maintained. This experiment provides a research idea for the fast detection of prohibited articles on mobile devices.

1. INTRODUCTION

Security inspection of prohibited articles as an important technical means to maintain the safety of public transportation, is widely used in customs, subways, airports and other public places, and is an important application field based on machine vision target detection. In recent years, with the great progress of deep learning in image classification, prohibited articles detection based on deep learning methods has gradually became the research focus and hotspot [1]. However, the security inspection equipment used in major airports, subways and other security inspection places is large in size and poor in mobility, which is not convenient for random inspection of luggage and parcels. Therefore, research on lightweight prohibited articles detection methods suitable for embedded mobile devices is of great significance for improving the flexibility of security inspections.

2. LIGHTWEIGHT CONVOLUTIONAL NEURAL NETWORK

The research of lightweight convolutional neural network has very important significance in practical application [2]. At present, the deep learning target detection methods are mainly divided into two categories, the two-stage series of algorithms based on region recommendations (R-CNN, Fast-RCNN, Faster-RCNN, R-FCN, etc.) and the single-stage series of algorithms based on regression problems (YOLO, SSD, etc.). These algorithms have advantages in detection precision, but due to the complex network model, the large scale of calculation, and the limited hardware resources of the embedded platform, it is difficult to apply to the mobile platform. In contrast, the lightweight target detection
network is lighter and more efficient, while maintaining a high-precision network model [3]. Therefore, the use of lightweight convolutional neural networks for prohibited articles detection has important engineering practical application significance.

2.1. MobileNet-SSD Network

The MobileNet-SSD target detection algorithm is an algorithm that uses MobileNet as the backbone of the network and uses the MobileNet neural network structure to replace the VGG16 convolutional neural network structure in the traditional SSD algorithm to extract features and combine it with SSD target detection. The SSD algorithm is based on the convolutional neural network structure of VGG16 and adds 4 convolutional layers to extract feature information [4]. It follows the YOLO target detection idea—completing the regression frame and classification at one time, and then joins the Faster R-CNN, using the anchor to improve recognition precision. The MobileNet-SSD target detection algorithm combines the advantages of MobileNet and SSD to ensure the precision of target detection under the premise of small calculation amount and fast measurement speed. The network structure of MobileNet-SSD is shown in Figure 1.

2.2. YOLOv3-Tiny Network

YOLOv3-tiny is a simplified model of YOLO. It consists of 9 convolutional layers and 6 pooling layers. Its size is only about half of the standard YOLO model. Although the recognition precision is slightly lower, the detection speed is far much faster than the standard YOLO model. The backbone network of YOLOv3-Tiny mainly has 7 convolutional layers with a size of 3×3 and 6 pooling layers. Its step size of the first 5 pooling layers is 2 while the step size of the last layer is 1, and the output size of the overall network is 13×13. Because the image size of the input convolutional neural network is different, the image size of the output is also different. The YOLOv3-Tiny network consists of a shallow network and a deep network. The deep network with a shallow convolutional layer is easier to characterize small target objects. On the contrary, the deep network with a deep convolutional layer is easier to characterize large target objects. This kind of network structure makes YOLOv3-Tiny have good detection performance for targets of different sizes [5]. The network structure of YOLOv3-Tiny is shown in Figure 2.
3. DATASETS AND EVALUATION INDICATORS

3.1. Experimental Datasets
Dataset is an important part of deep learning network. The applicability of dataset greatly affects the performance test effect of deep learning network and its feasibility in practical application. On the one hand, before the appearance of convolutional neural network, the detection of prohibited articles in public places was based on the traditional algorithm, and the traditional detection method did not require amounts of security image data, so there was no unified and universal dataset for research. On the other hand, deep learning algorithm has been applied in the field of security inspection for a short time, and most public security inspection images involve citizen privacy issues. Therefore, there are few public X-ray image datasets for security inspection. Common security inspection image datasets mainly include GDXray and SIXray (security inspection X-ray). The characteristics of the two datasets are compared, as shown at Table 1.

| Datasets | Number of security images | Number of images with prohibited Articles | Prohibited Articles | Characteristic |
|----------|---------------------------|------------------------------------------|--------------------|---------------|
| GDXray   | 8150                      | 1223                                     | Guns, grenades, darts, blades | The data set is small, the image background is simple, and the image is gray image. Compared with the commonly used pseudo color image, it lacks very important color information. |
| SIXray   | 1059231                   | 8929                                     | Guns, knives, pliers, wrenches, hammers and scissors | The number of images in the data set is large, the pseudo color image is closer to the actual security image, and the image background is complex, which is more challenging and meaningful |

SIXray dataset has a large number of images, which contain pseudo color images closer to the actual security image, and the image background is complex, so it is more challenging and meaningful. Therefore, this paper uses SIXray dataset to detect prohibited articles. A total of 8929 images of prohibited articles were used to divide the dataset. The test set and training set accounted for 80%, and the verification set accounted for 20%.
3.2. Evaluation index of the model
Average precision (AP) and mean average precision (mAP) are commonly used in the field of target
detection as the evaluation index of model detection effect. AP means the average value of the model's
detection precision for a certain type of object on the test set, and mAP means the average value of all
AP. The calculation formula is shown in (1) ~ (4):

\[ \text{Precision} = \frac{TP}{TP + FP} \]  

(1)

\[ \text{Recall} = \frac{TP}{TP + FN} \]  

(2)

\[ AP = \int_0^1 P(R) dr \]  

(3)

\[ mAP = \frac{1}{N} \sum_{i=0}^{N-1} AP_i \]  

(4)

TP means the number of positive samples that the model detects correctly; FP means the number of
positive samples for detecting errors; FN mean the number of negative samples for detecting errors;
Precision and recall represent precision and recall respectively; P(R) is a function with recall as
independent variable and precision as dependent variable; N means the number of objects. There are five
kinds of prohibited articles in the dataset used in this paper. Detection speed is one of the important
indexes to evaluate the engineering practicability of the model. The detection speed of the model is related
to the complexity of the target detection network, the depth of the network, and the computing power of
the hardware. Different object detection algorithms often have different map and detection speed. Object
detection algorithms in the detection of prohibited articles should not only meet certain detection
precision, but also ensure that the detection speed of network model on the security equipment meets the
actual requirements of the project. In this paper, the average detection time of a single frame image is
used to evaluate the detection speed of the network.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1. Experimental Equipment and Model Parameters
In this experiment, the system used to train the network model is Windows 64 bit, and the graphics card
is GTX 1660s. The mobile device adopts Raspberry Pi 4B microprocessor, CPU: ARM Cortex-A72
1.5GHz, memory: 8g, operating system: Linux Raspberry Pi 5.4.51. In this experiment, the random
gradient descent method is used to optimize the parameters, in which the parameters are set as follows:
the decay coefficient is set to 0.0001, the momentum parameter is set to 0.9, and the initial learning rate
is set to 0.001, and the number of iterations is 120. The loss function in the process of MobileNet-SSD is
shown in Figure 3 and the loss function in the process of YOLOv3-Tiny is shown in Figure 4.

Figure 3. Change in network loss function of MobileNet-SSD
4.2. Experimental Results

In this paper, a comparative experiment is carried out on the public SIXray dataset. There are five kinds of detection targets—guns, knife, pliers, wrench and scissors. YOLO, YOLOv3-Tiny and MobileNet-SSD convolutional neural network models are used for prohibited articles detection training. The trained model is transferred to Raspberry Pi mobile device for prohibited articles detection and the detection speed of the network model is tested.

Table 2 shows the detection results of three networks on Raspberry Pi mobile device. It can be seen from the table that the detection precision of YOLO is the highest, and the detection precision of prohibited articles of YOLOv3-Tiny is significantly higher than that of MobileNet-SSD. In addition, since both MobileNet-SSD and YOLOv3-Tiny are lightweight networks, their detection speeds are similar and about 10 times faster than YOLOv3 detection speeds. The image detection result of YOLO-Tiny is shown in Figure 5.

Table 2: Comparison of Detection Precision and Detection Speed of the Model

| Detection model | Average precision of Gun | Average precision of Knives | Average precision of Wrench | Average precision of Pliers | Average precision of Scissors | Mean average precision | Average detection speed (FPS/s) |
|-----------------|--------------------------|-----------------------------|-----------------------------|-----------------------------|-------------------------------|------------------------|--------------------------------|
| YOLOv3          | 90.50%                   | 77.97%                      | 79.61%                      | 85.94%                      | 79.66%                        | 82.74 %                | 0.21                           |
| MobileNet-SSD   | 91.37%                   | 68.11%                      | 43.47%                      | 54.42%                      | 48.07%                        | 61.09%                 | 2.12                           |
| YOLOv3-Tiny     | 90.11%                   | 75.72%                      | 72.80%                      | 77.80%                      | 74.07%                        | 78.10%                 | 2.03                           |

Figure 4. Change in network loss function of YOLOv3-Tiny

Figure 5. Detection result of YOLOv3-Tiny
4.3. Conclusion
In this paper, YOLO, YOLOv3-Tiny and MobileNet-SSD convolutional neural networks are used to detect prohibited articles in Raspberry Pi 4B mobile device. The experimental results are shown as follows.

4.3.1 In terms of detection precision, YOLO model has the highest average detection precision, which is 82.74%. YOLO tiny’s detection precision is slightly lower than YOLO neural network, but the decline is not significant, which is 78.10%; MobileNet-SSD has a low network level, and the network focuses on multi-scale target detection, while the size of prohibited articles in SIXray dataset is similar, so the detection precision is low.

4.3.2 In terms of detection speed, the detection speed of YOLO in mobile platform is very low, only 0.21 FPS / s, due to its deep network level, large number of parameters and large network model. The detection speed of YOLOv3-Tiny and MobileNet-SSD lightweight network model is about 10 times of YOLO detection speed, which greatly improves the detection speed, proving the feasibility of the deployment of prohibited articles detection in mobile platform.

4.3.3 The current research on the detection of prohibited articles mostly focuses on the use of complex detection models to improve detection precision. However, this article not only detects the detection precision of the proposed network on mobile devices, but also detects its detection speed, which provides a reference for the practical application of prohibited articles detection technology on mobile devices.

4.3.4 Although the use of lightweight models can greatly increase the detection speed, the detection speed on the Raspberry Pi mobile device is still far from the actual engineering application requirements. It will be a meaningful attempt to improve the computing power of the Raspberry Pi, or use other mobile devices with stronger computing performance. In addition, improving the structure of the lightweight model based on the characteristics of security images to improve detection precision and detection speed will be one of the important research directions in the future.

ACKNOWLEDGMENT
This research is financially supported by the Guangzhou Municipal Special Fund Project for Scientific and Technological Innovation and Development (202002020068, 202002030147).

REFERENCES
[1] Saavedra D, Banerjee S, Mery D. Detection of threat objects in baggage inspection with X-ray images using deep learning[J]. Neural Computing and Applications, 2020: 1-17.
[2] Xun Hu, Hong Li, Xinrong Li, Chiyu Wang. MobileNet-SSD MicroScope using adaptive error correction algorithm: real-time detection of license plates on mobile devices[J]. IET Intelligent Transport Systems,2020,14(2).
[3] Qi Hangyu, Xu Tianhua, Wang Guang, Cheng Yu, Chen Cong. MYOLOv3-Tiny: A new convolutional neural network architecture for real-time detection of track fasteners[J]. Computers in Industry,2020,123.
[4] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C] // European conference on computer vision. Springer, Cham, 2016: 21-37.
[5] Wentao Li, Yan Zhang, Jinqiu Mo, Yanming Li, Chengliang Liu. Detection of Pedestrian and Agricultural Vehicles in Field Based on Improved YOLOv3 tiny[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(S1):1-8+33.