PAPER • OPEN ACCESS

Prohibited Items Detection in X-ray Images Based on Attention Mechanism

View the article online for updates and enhancements.

Prohibited Items Detection in X-ray Images Based on Attention Mechanism

Tianfen Liang[1, a], Bo Lv[2], Nanfeng Zhang[3], Jinhao Yuan[4], Yanxi Zhang[5], Xiangdong Gao[6]

[1]Guangdong Provincial Welding Engineering Technology Research Center, Guangdong University of Technology Guangzhou, China

[2]Guangdong Provincial Welding Engineering Technology Research Center, Guangdong University of Technology Guangzhou, China

[3]Huangpu Customs Technical Center Dongguan, China

[4]Guangdong Provincial Welding Engineering Technology Research Center, Guangdong University of TechnologyGuangzhou, China

[5]Guangdong Provincial Welding Engineering Technology Research Center, Guangdong University of Technology Guangzhou, China

[6]Guangdong Provincial Welding Engineering Technology Research Center, Guangdong University of Technology Guangzhou, China

[a]13050592889@163.com

[*]Corresponding author: gaoxd@gdut.edu.cn

Abstract. Security inspection is an important measure to ensure public safety. At present, X-ray security inspection equipment is widely used in security checkpoints. However, it is not efficient to recognize prohibited items in X-ray images manually. Automated security inspection system has become the development trend of security field. In this paper, the Yolov4 detector was used to detect prohibited items in X-ray images. In order to improve the detection performance, we added CBAM attention modules to different parts of Yolov4. A public dataset is used for simulation experiments, which shows that the addition of CBAM can effectively improve the detection performance of the detector.

1. Introduction

X-ray security inspection is a mature security technology, which is widely used in security checkpoints. X-ray images can show objects hidden in luggage, so prohibited items can be detected by X-ray images. At present, security inspection requires human operators to recognize prohibited items in X-ray images. Manual detection of prohibited items is not efficient, especially during peak periods of movement. In the need of a large number of security inspection, the automated security inspection system has become the development trend of security field. The imaging principle of X-ray image is different from the imaging principle of optical image, so X-ray image has different characteristics. X-ray can pass through different objects, causing the contour of each object to overlap in the X-ray image. In addition, X-ray image can easily obtain the contour of the object, but the detailed features of the object surface are easy to lose. Therefore, it is a difficult task for detecting prohibited items in X-ray images. Now, the study of

automated security is mostly based on deep learning model. Mery et al. [1] used deep learning model to recognize objects in X-ray images, and the results showed that it is possible to use deep learning models to design automated security devices. Akcay et al. [2] conducted a large number of experiments to compare the performance of traditional machine learning and deep learning in processing X-ray images, and they used the deep learning model to recognize prohibited items. The results show that the application of deep learning to the detection of prohibited items in X-ray images can well complete the task of classification and object detection. The classification task simply classifies the objects, which cannot meet the requirements of automated security inspection system. Automated security inspection system needs to find prohibited items and its position in the X-ray image, so the object detection algorithm is more widely used in the field of automated security inspection. Zhang et al. [3] improved the SSD by adding modules such as asymmetrical tiny convolution module, dilated convolution multi-view module and fusion strategy of multi-scale feature map to enhance the detector 's learning of overlapped objects in the global field of view. Zhang et al. [4] improved FSSD and extracted low-level features with strong semantic information by using semantically rich modules composed of dilated convolution. In addition, they added residual modules with residual blocks into FSSD to extract features. This paper proposes an object detection detector based on Yolov4 [5], in which attention modules are added, and a comparative experiment is conducted to study the influence of attention modules on the detection performance of prohibited items in X-ray images.

2. Methods

2.1. Yolov4
Yolov4 is an excellent one-stage detectors in the Yolo series. The one-stage detectors realize the end-to-end detection. The biggest advantage of these detectors is that it is fast enough to real-time detection. Yolov4 does not need to generate region proposals and can directly predict the object's bounding boxes and categories by simply inputting images into the convolutional neural network. The detector divides the picture into $N \times N$ grids, and each grid is responsible for the object that falls on the grid, which is the basic idea of the Yolo detectors. Yolov4 used a large number of optimizations means to improve the detection performance. Yolov4 used the complex CSPDarkNet53 as the backbone network, which had a large receptive field and large parameters. Yolov4 fused 4 images at a time to enrich the background of training images by using mosaic data enhancement method. The Yolov4 structure is shown in Figure 1. The input is extracted through the backbone network, and then the three groups of feature maps of different sizes are fused to get the output head.

2.2. CBAM
CBAM [6] can effectively improve the detection performance of deep learning model, its structure is very simple and the number of its parameters is very small. The attention mechanism scans the entire image to find noteworthy regions and channels, and then invests more resources in the corresponding regions and channels. The mechanism of rapidly selecting the effective information from a large amount of information and suppressing the invalid information are very similar to human vision. Therefore, adding attention module to deep learning model can effectively improve the detection performance of the model. There are two sub-modules of CBAM: channel attention module and spatial attention module. The channel attention module first performs average-pooling and max-pooling operations on input F to extract its channel information, and then generates the corresponding channel attention map through a multi-layer perceptron (MLP). Finally, the two channel attention maps were summed and the sigmoid activation function was used to generate the channel attention *Mc (F)*. Therefore, the formula for calculating channel attention is:

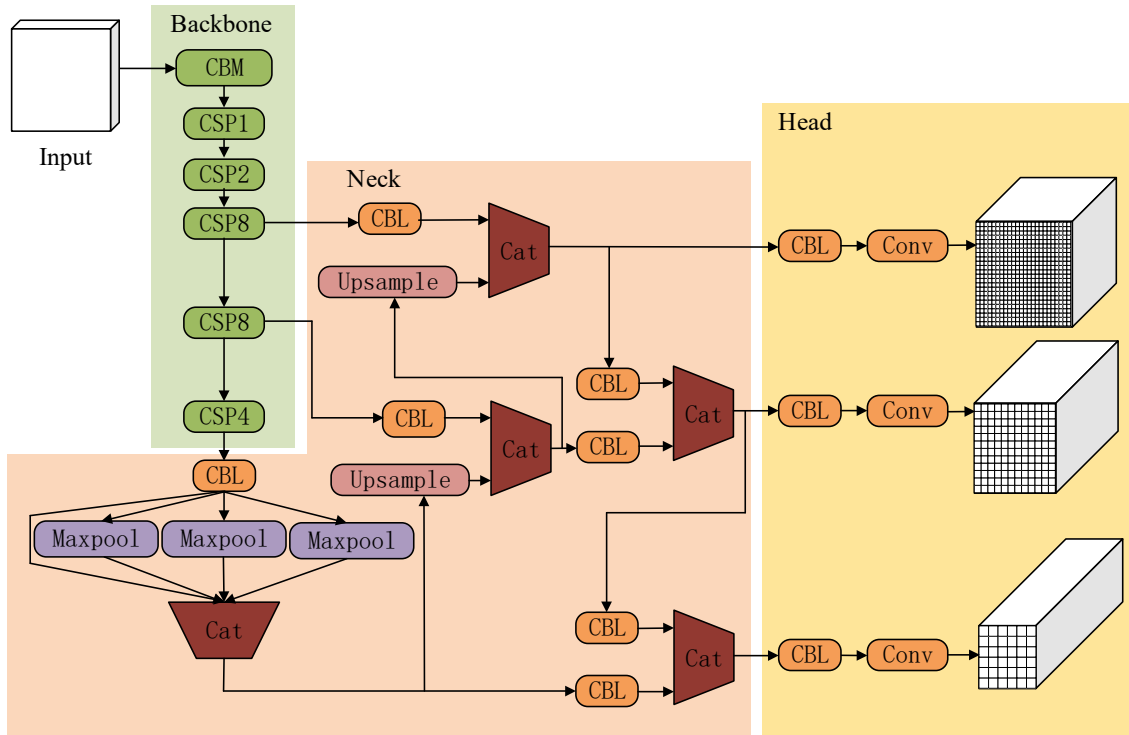$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \tag{1}$$
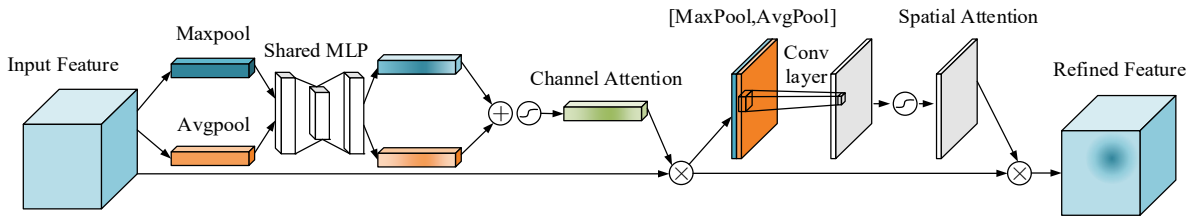
Figure 1 The diagram of Yolov4

Figure 2 The diagram of CBAM

The spatial attention module firstly carries out average-pooling and max-pooling on the input *F* along the channel dimension, and then uses dilated convolution to extract efficient aggregated spatial context information. Finally, the sigmoid activation function is used to generate spatial attention *Ms (F)*. Its formula is expressed as follows:

$$M_s(F) = \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)])) \tag{2}$$

CBAM effectively combines channel attention and spatial attention in order, as shown in Figure 2.

2.3. Yolov4-CBAM

X-ray can penetrate different objects. When multiple objects are stacked together, X-ray will pass through multiple objects. The contour of each object is overlapped, so the X-ray image has complex contour information. In order to capture important features from complex X-ray images and improve the performance of detector, spatial attention and channel attention
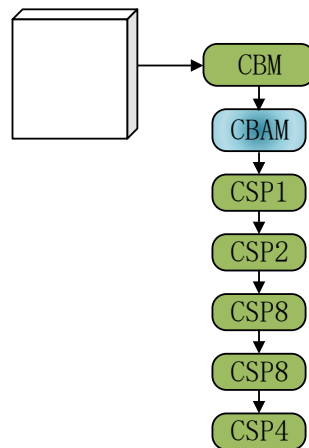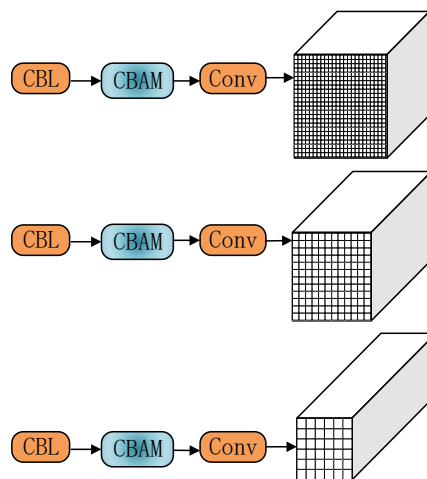
Figure 3 Attention model integrated into backbone

Figure 4 Attention model integrated into head

Table 1   Data partition

Data	Gun	Knife	Pliers	Scissors	Wrench
Train	3498	2126	3787	811	2144
Test	1480	953	1583	321	940
Total	4978	3079	5370	1132	3084

were integrated into Yolov4. The CBAM module was added into the backbone network CSPDarknet53 to improve the capability of feature extraction. We added CBAM to the backbone network, as shown in Figure 3. Yolov4 head outputs detection results of three different scales, which store the information about objects detected by the detector. We added a CBAM module in front of each head of Yolov4, as shown in Figure 4. We only added 4 CBAM modules into Yolov4. The weight of the original model was 250.50m, and the weight of the model was 250.53m after adding the CBAM modules, so the increase of parameters was almost negligible.

3. Experiment and result

3.1. Experimental Setup

At present, only GDXray [7] and SIXray [8] can be used in the public data sets for prohibited items detection. GDXray is the first publicly available large data set containing X-ray images of prohibited items, including handguns, shuriken and razor blades. The images in GDXray data set are grayscale images, in which the contours of objects are clear and easy to distinguish, most of the images have simple backgrounds, and few objects overlap together. The SIXray dataset provides a large number of X-ray images with six categories of objects: gun, knife, wrench, pliers, scissors and hammer. Objects in SIXray are randomly placed, and the background of the images is complex and many objects are seriously overlapping. This data set is suitable for the comparative experiment. Therefore, this paper uses Yolov4 and Yolov4-CBAM proposed in this paper to recognize prohibited items in SIXray. Since a handful of images contain hammers, only 8,908 images of SIXray were selected for the experiment, in which there are five types of objects: gun, knife, scissors, pliers and wrench. We use 70% of the images as the training set and 30% of the images as the test set. The number of objects in each data set is shown in Table I. All of the models were trained by using GTX 1080Ti in the PyTorch framework, and the Adam optimizer was used for optimization. Training parameters are as follows: initial learning rate (0.001), Weight-decay (5e-4), batch size (4).

3.2. Result and analysis

MAP was used as the evaluation index, which considered the recall and the precision of the testing process. Transfer

Table 2   Comparison of improved Yolov4 mAP

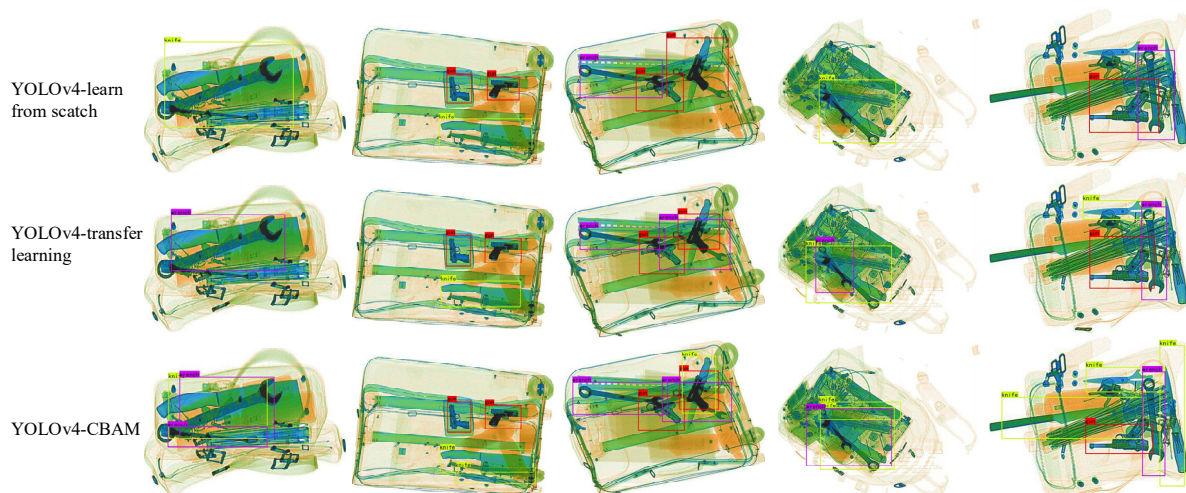| Detector	Gun AP	Knife AP	Pliers AP	Scissors AP	Wrench AP	mAP
Yolov4-learn from scratch	91.90	64.34	76.97	69.76	56.94	71.98
Yolov4-transfer learning	94.99	75.56	86.41	81.38	78.60	83.39
Yolov4-CBAM	95.38	80.06	87.16	81.56	83.39	85.51

Figure 5 The detection effect of each detector

learning can make the model learn quickly, so the comparison between learning from scratch and transfer learning is added in the experiment. It can be seen from Table II that the mAP of learning from scratch is 71.98%, while the mAP of using transfer learning is 83.39%, indicating that transfer learning is

effective in this experiment. The AP of all prohibited items increased with the addition of CBAM in Yolov4. The knife AP increased 4.5%, the wrench AP increased 4.79%, and the mAP increased 2.12%. These improvements indicate that CBAM can effectively improve the detection performance of the detector. According to Table II, gun AP is the highest, reaching 95.38%. Pilers AP also did well, reaching 87.16%. It must also be mentioned that knife AP is the lowest. Because some of the knives overlap in the X-ray image, the detector does not recognize all of them. Scissors is a small object, which takes a small proportion in the X-ray image, so it is easy to ignore. Figure 4 shows that most prohibited items can be detected. But the objects are stacked randomly, and the background is complex. Some hidden objects are not recognized and overlapping objects may be ignored, which is the problem faced by automated security.

4. Conclusion

In this paper, we take Yolov4 as the basic model, and then add attention modules to different parts of the model as prohibited items detectors. Adding the CBAM module has almost negligible effect on the parameters of the detector. We conducted a comparative experiment with SIXray data set to analyze the effect of adding attention modules on the model performance. The result shows that the addition of CBAM modules can effectively improve the detection performance of the detector for all kinds of prohibited items. We have successfully applied deep learning model to automated security inspection. There are still some shortcomings, such as overlapping objects and small objects are easy to ignore. In the next step, we will improve the network structure to solve these problems and improve the detection performance of automated security.

References

[1]   Mery D., Svec E., Arias M., et al. (2017). Modern computer vision techniques for x-ray testing in baggage inspection. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 47(4), 682-692.

[2]   Akcay S., Kundegorski M.E., Willcocks C.G., et al. (2018). Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery. IEEE transactions on information forensics and security, 13(9), 2203-2215.

[3]   Zhang, Y.K., Su Z. G., Zhang H. G. et al. (2020). Multi-scale Prohibited Item Detection in X-ray Security Image. Journal of Signal Processing, 2020, 36(07): 1096-1106.

[4]   Zhang, Y.T., Zhang, H.G., Zhao, T.F., et al. (2020). Automatic detection of prohibited items with small size in X-ray images. Optoelectronics Letters, 16(4), 313-317.

[5]   Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[6]   Woo, S., Park, J., Lee, J. Y., et al. (2018). CBAM: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). Munich. pp. 3-19.

[7]   Mery, D., Riffo, V., Zscherpel, U., et al. (2015). GDXray: The database of X-ray images for nondestructive testing. Journal of Nondestructive Evaluation, 34(4), 1-12.

[8]   Miao, C., Xie, L., Wan, F., et al. (2019). SIXray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach.   pp. 2119-2128.