ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20220222034206
Original scientific paper

Intelligent Detection of Dangerous Goods in Security Inspection Based on Cascade Cross Stage YOLOv3 Model

Jianjun WU*, Shaowen LIAO

Abstract: At present, it mainly depends on the human eye to identify the X-ray scanning image, when the security detector is used to detect the dangerous goods in the baggage. It is labor intensive and prone to false or missed detection. This paper proposes an intelligent detection method of dangerous goods in security inspection based on a novel cascaded cross-stage YOLOv3 model (abbreviated to CCS-YOLOv3). Considering the different sizes, disorderly lay or serious overlap of various objects in the scanning image, this method first enhances the scanned image to improve the quality of the data set. After that, the traditional YOLOv3 is improved by cascading cross-stage mode, and the backbone network of YOLOv3 is improved to cascade cross-stage Darknet. And then the backbone network is followed by a spatial pyramid pooling (SPP) module. Following that, the feature pyramid network (FPN) is connected in series with a bottom-up feature pyramid structure to realize the feature fusion. The results of model Ablation experiment and baggage scanning image detection show that the cascade cross-stage YOLOv3 model significantly improves the image detection speed and precision, and the model is effective and feasible.

Keywords: cascade cross stage networks; detection of dangerous goods; feature fusion; intelligent security inspection; YOLOv3 model

1 INTRODUCTION

With the increasing travel flow of public transport, in order to ensure the travel safety of passengers, rapid and accurate identification of dangerous goods is an important problem to be solved by security inspection [1]. At present, the security inspection process is still highly dependent on manual work. Security inspectors use experience to observe whether the X-ray scanning image contains dangerous goods in the baggage. This is time consuming and labor intensive, and it is prone to false and missed detection. In addition, in order to ensure detection accuracy, reducing the speed of the belt of the security inspection machine is frequently used, which is prone to causing crowd congestion. Therefore, how to realize intelligent assisted security inspection is an urgent task in public transport.

Traditional image processing technology was used to enhance the image quality of X-ray scanning in the earlier period. Abidi et al. [2] reduced the noise in security inspection images by fusing low-energy and high-energy X-ray images and using background difference. In 2012, Yang Xiaogang et al. [3] used bilateral filtering to remove X-ray image noise while preserving edge information, and improved image contrast by adding constraints in local histogram equalization. With the development of machine learning, many scholars have begun to explore the use of machine learning methods for automatic detection of dangerous goods in security inspections. In 2011, Baştan et al. [4] applied the BoVW method to the detection of security inspection images for the first time, and classified contraband on a data set containing 200 X-ray security inspection images with an average detection accuracy of 65%. Franzel et al. [5] proposed a sliding window detection method in 2012 that uses a linear support vector machine (SVM) and a histogram of oriented gradients (HoG) [6] simultaneously in the sliding window, by fusing multiple viewpoints to find the true detection intersection. In 2013, Turcsany et al. [7] proposed using SURF [8] feature detector and feature descriptor based on BoVW method to train SVM [9], and obtained 99.07% accuracy and 4.31% error rate. In 2015, Baştan et al. [10] applied branch definition algorithm and SVM to evaluate various artificial features and description operators in about 6000 X-ray security inspection sets. In 2015, Mery et al. [11] proposed a multi-stage general method. This method first extracts features using a feature descriptor and a nearest neighbour classifier, then matches the key points of continuous images from different perspectives, and finally analyzes the key point matching of two continuous images in multiple views. Traditional machine learning-based detection relies heavily on manually selected low-level features. Therefore, the objects could not be well detected if these features have insufficient robustness and generalization in the case of changeable scenes.

In recent years, with the rapid development of computing hardware and the innovation of deep learning theory, the detection based on deep learning has been successfully applied to the detection of dangerous goods, which effectively improves the detection accuracy and speed compared with the traditional methods. In 2017, Akçay [12] first applied the deep learning model to the detection of dangerous goods in X-ray images, and used the models based on sliding window convolution neural network (CNN), Faster-RCNN [13] and R-FCN [14] to detect guns and other dangerous goods on DBF2/6 data set. In 2018, Liang et al. [15] explored the performance of Faster-RCNN, R-FCN and SSD (Single Shot MultiBox Detector) models [16] in single view and multi view X-ray images, and further improved the accuracy of dangerous goods through the detection of combined multi view objects. Subsequently, Liang et al. continued to train and evaluate SSD and Faster-RCNN [17] on data sets containing more types of dangerous goods, and further validated Faster-RCNN's performance in dangerous goods detection. On the other hand, the classified detection model Yolo [18] has received widespread attention and has been continuously improved since its introduction [19]. Its detection has strong robustness and fast detection speed, making it more suitable for dangerous goods detection. Liu et al. used YOLOv2 [20] model to train in SASC data set, and achieved 94.5% detection accuracy. Dhiraj and others [21] also use YOLOv2 and Tiny YOLO [22] models to detect contraband objects such as guns, razor blades and knives from baggage X-ray images. In addition, in-depth learning models such as adversarial learning are also used for dangerous goods detection in baggage [23-25]. Most of researches make use of single dangerous goods or black-and-white images, but Miao et al. created a data set SIXRay [26] in a real-world scene and proposed a deep class balanced hierarchical refinement (CHR) framework to detect suspicious goods, which will be of great assistance to people's research in this field.

Although YOLOv3 has the characteristics of strong robustness and fast detection speed, the accuracy and recall of the model in actual detection are still insufficient. In addition, the accuracy of the model for small object detection is not satisfactory. Considering the drawbacks of YOLOv3 and the above traditional methods, an intelligent detection method of dangerous goods based on cascaded cross-stage YOLOv3 model (CCS- YOLOv3) is proposed. In the new model, traditional YOLOv3 is improved by cascading cross-stage model, and the backbone network of YOLOv3 is improved to cascade cross-stage Darknet. The backbone network is followed by a spatial pyramid pooling (SPP) module, which then implements feature fusion using the feature pyramid network (FPN) and bottom-up pyramid series connection. It is anticipated that the proposed method can obtain good detection performance.

2 TRADITIONAL YOLO V3 MODEL FOR OBJECT DETECTION: AN OVERVIEW

YOLOv3 is an improved end-to-end detection model. While retaining the fast calculation speed of YOLO, it has also achieved good results in small object detection. The topology of YOLOv3 is shown in Fig. 1. Firstly, take the image to be detected as the input, pass through a CBL structure composed of convolution layer, BN layer and Leaky ReLU activation function, and pass through the Darknet53 backbone network with the fully connected layer removed. Inspired from the structural idea of ResNet, the Darknet53 backbone network has five residual blocks. Each residual block is composed of a CBL structure and several residual units. The structure of the residual unit is two CBL structures and connected by residual. Next, the FPN module is added to the residual block structure, and then three different scales of output are output through CBL and convolution layer.

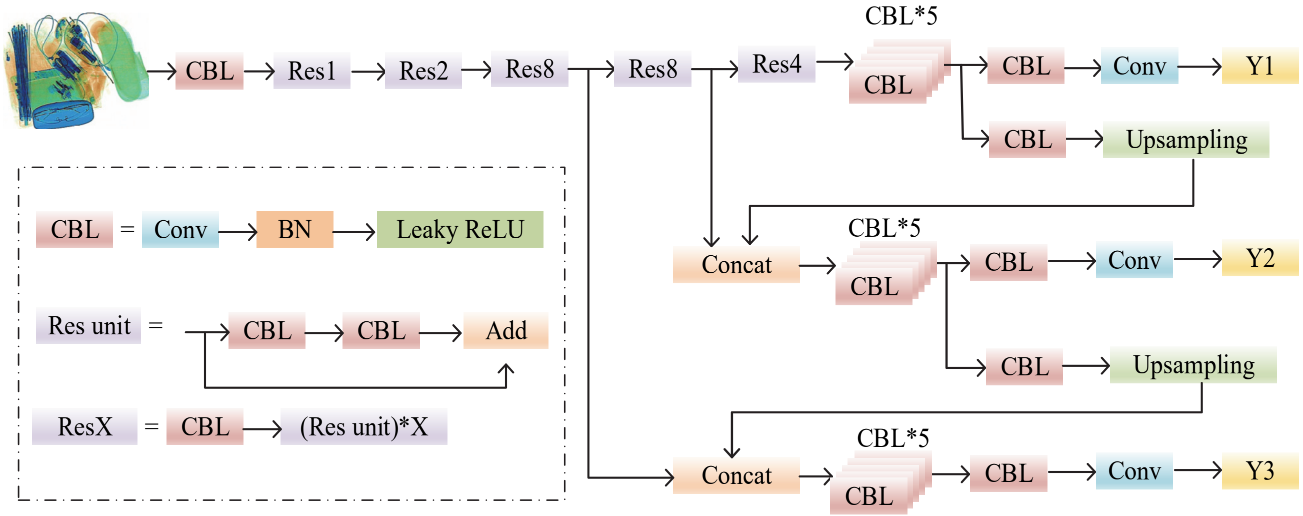


Figure 1 Topological diagram of the YOLOv3 model

As shown in Fig. 1, firstly, YOLOv3 divides the input image into multiple grids in horizontal and vertical directions based on the sliding window algorithm. If the center point of the object to be detected falls in a grid, the grid is responsible for the prediction of the object to be detected. Secondly, in each grid, predict the confidence of a certain number and size of anchor frames and the objects to be detected in the anchor frames, and further filter out the redundant anchor frames after filtering out the low confidence anchor frames. Finally, the category of the remaining anchor frames is predicted.

In terms of backbone network, Darknet53 solves the problem of gradient explosion when deepening the network depth by introducing the idea of residual network with cross layer connection, improves the feature extraction ability, reduces the computational complexity and improves the running speed of the model. In the aspect of feature extraction, the model uses FPN for multi-scale prediction to enhance the ability of feature expression. FPN structure adopts top-down and horizontal connection to integrate the features of adjacent scales, and output three different scale feature maps, which are respectively responsible for detecting large, medium and small objects. By fusing the deep and shallow abstract semantic features and detail features, it provides semantic guidance for the shallow features and increases the semantics of the shallow features, which is conducive to the detection of small objects in the image.

In terms of anchor frame design, YOLOv3 model resets the anchor frame size with the help of K-means clustering, which can be divided into multiple groups and multiple in each group. Small anchor frames are applied in feature maps with high resolution. In terms of loss function, in order to detect multiple objects in the same image, YOLOv3 uses Logistic classifier instead of Softmax. The loss function is expressed by cross entropy function, and its definition is shown in Eq. (1):

$$\begin{aligned}
Loss = {} & \lambda_{coord}\sum_{i=0}^{M^2}\sum_{j=0}^{N} I_{ij}^{obj}[(x_i-\hat{x}_i^j)^2+(y_i-\hat{y}_i^j)^2]+\\
& \lambda_{coord}\sum_{i=0}^{M^2}\sum_{j=0}^{N} I_{ij}^{obj}[(\sqrt{w_i^j}-\sqrt{\hat{w}_i^j})^2+(\sqrt{h_i^j}-\sqrt{\hat{h}_i^j})^2]-\\
& \sum_{i=0}^{M^2}\sum_{j=0}^{N} I_{ij}^{obj}[\hat{C}_i^j\log(C_i^j)+(1-\hat{C}_i^j)\log(1-C_i^j)]- \qquad (1)\\
& \lambda_{noobj}\sum_{i=0}^{M^2}\sum_{j=0}^{N} I_{ij}^{noobj}[\hat{C}_i^j\log(C_i^j)+(1-\hat{C}_i^j)\log(1-C_i^j)]-\\
& \sum_{i=0}^{M^2} I_{ij}^{obj}\sum_{c\in class}[\hat{P}_i^j\log(P_i^j)+(1-\hat{P}_i^j)\log(1-P_i^j)]
\end{aligned}$$

Among them, the input image will be divided into grids by the program, and N anchor frames will be generated for each grid. After each anchor frame is processed by the program, the corresponding bounding boxes will be obtained, and a total of bounding boxes will be obtained. Next, use the loss function to calculate and update the weight.

Parameter I_{ij}^{obj} : indicates whether the j-th a priori box of the i-th grid is responsible for the target object. If it is responsible, I_{ij}^{obj} = 1, otherwise it is 0.

Parameter confidence $\hat{C}_i^j$: represents the real value. The value of $\hat{C}_i^j$ is determined by whether the bounding box of the grid is responsible for predicting an object. $\hat{C}_i^j$ = 1, otherwise it is 0.

3 CASCADE CROSS STAGE YOLOV3 MODEL

Cascade cross stage YOLOv3 model (CCS-YOLO v3) has three improvements over the traditional YOLOv3 model. The first is to improve the backbone network of YOLOv3 into a cascade cross-level Darknet network by using cascade cross-level method. The second is to add a spatial pyramid pooling (SPP) module after the backbone network. The third is to realize feature fusion by using FPN and bottom-up feature pyramid in series. The topology of CCS-YOLOv3 model is shown in Fig. 2.

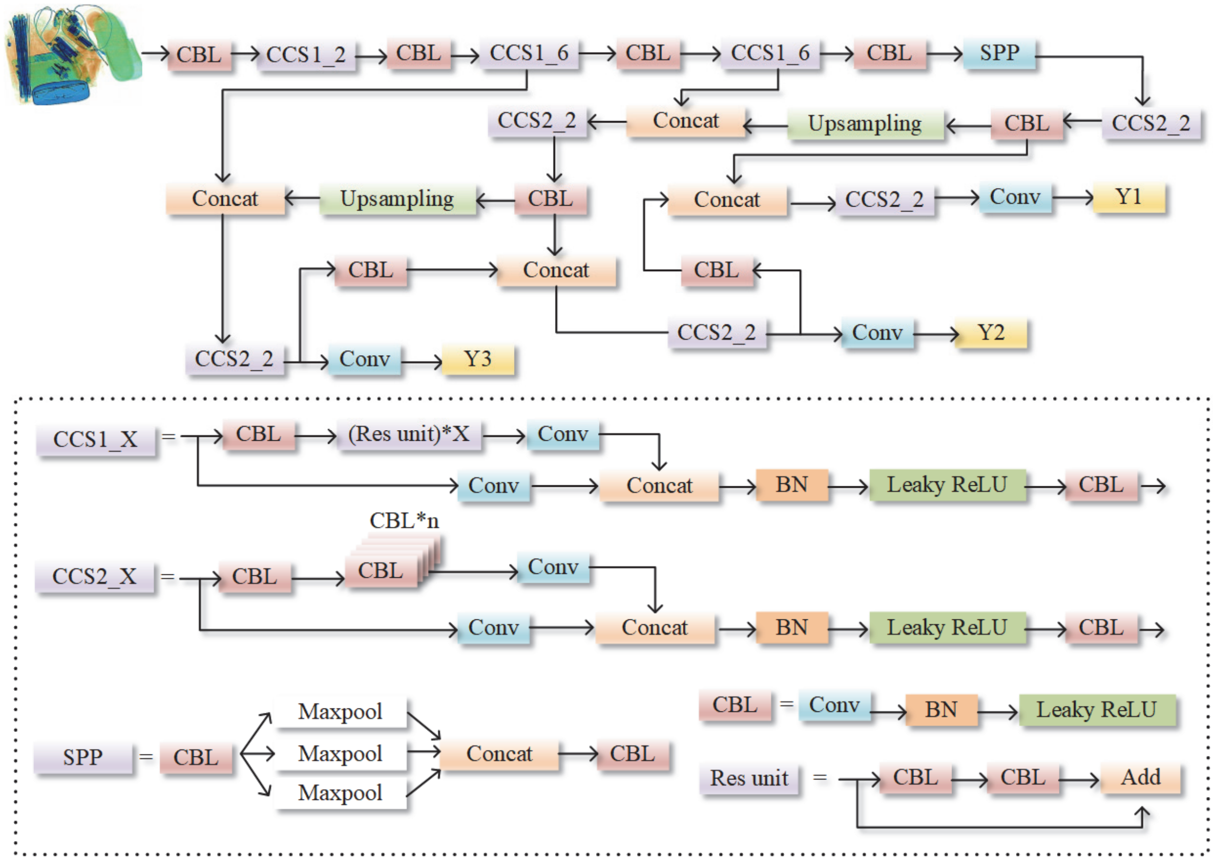


Figure 2 Topological diagram of the CCS-YOLOv3 model

As seen in Fig. 2, the CCS-YOLOv3 model first takes the image to be detected as the input, passes through a CBL structure composed of convolution layer, BN layer and Leaky ReLU activation function, and then passes through CBL and CCS1_6, CBL, CCS1_6, CBL structure. Secondly, by referring to the structural idea of residual network ResNet, the channel fusion is carried out after the branch passes through a convolution layer. Each residual block is composed of BN layer, Leaky ReLU activation function, a CBL structure and a different number of residual units. Finally, based on the FPN module of YOLOv3, a bottom-up feature pyramid module is connected in series for feature fusion.

In Fig. 2, the structures of CCS1_X and CCS2_X are derived from the CCS topology. The CCS first divides the characteristics of the foundation layer into two parts, one part is represented by the cross stage hierarchy, and the other part is directly integrated with the output of the previous part. In this way, the number of gradient paths can be doubled through the split and merge design, and the computational bottleneck with high computational power can be removed. It reduces the amount of calculation and ensures the accuracy of calculation. As shown in Fig. 2, the CCS1_X structure first passes through a CBL structure, X (X is a positive integer) Res unit structure and a convolution layer, and then carries out channel fusion after the branch passes through a convolution layer based on the residual idea, and then passes through the BN layer, Leaky ReLU activation function and a CBL structure. The structure of CCS2_X is roughly the same as that of CCS1_X, except that X combined residual ideas are short circuited and added, and the structure part of Res unit is 2 × X CBL structures instead.

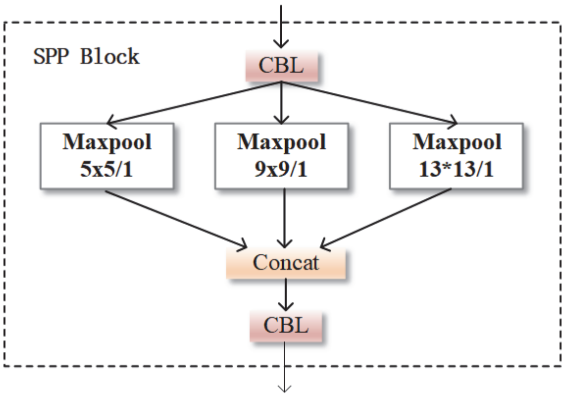


Figure 3 SPP Block

In addition, the SPP module in the CCS-YOLOv3 model adopts maxpool mode. Its structure is shown in Fig. 3. Here, the maxpool stride is 1, and the padding operation is used to make the output size of each layer of the pyramid the same. For example, 13 × 13 Input feature map, using 5 × 5 size pool kernel pooling, padding size is 2, so the feature map, after pooling is still 13 × 13 size. Considering that obtaining context information of different scales can increase the receptive field, and then enhance the whole feature system. Therefore, the SPP module is added after the backbone network of CCS-YOLOv3, which can effectively improve the acceptance range of backbone network features and make the separation of context features more significant.

Finally, the CCS-YOLOv3 model connects the bottom-up feature pyramid behind the FPN module to realize the feature fusion function. The basic structure of this part is based on the new path aggregation network of CCS2_X, which further improves the calculation speed and feature extraction of the model. In such a concatenation operation, the FPN module transmits strong semantic features from top to bottom, while the feature pyramid from bottom to top transmits strong location features. Therefore, the CCS-YOLOv3 model can further improve the feature extraction by changing the topology of feature fusion.

4 TEST RESULTS AND COMPARATIVE ANALYSIS OF DANGEROUS GOODS IN SECURITY INSPECTION

4.1 Experimental Environment Setting and Evaluation Index

The data set of X-ray scanning images used in the detection experiments came from the data set SIXray provided in the literature [26]. Some images in this data set are shown in Fig. 4, which are all composed of X-ray scanning images in the real scene of subway security inspection. The 7143 images from the data set SIXray were selected to form the learning sample set, and 1786 images were selected to form the test sample set. The two types of sample sets each contain 5 types of dangerous goods positive samples and 1 type of negative samples. The sample categories are knife, gun, scissors, wrench and pliers. It can be seen from the figures that the quantity of all kinds of dangerous goods is different. The pliers with the largest number are 4 times that of the scissors with the smallest number. Therefore, the sample size of all kinds of dangerous goods is not balanced. In addition, it can be seen from Fig. 4 that the volume of dangerous goods objects is small, scattered in X-ray images, and their positions are random. These characteristics also increase the difficulty of intelligent detection of dangerous goods.

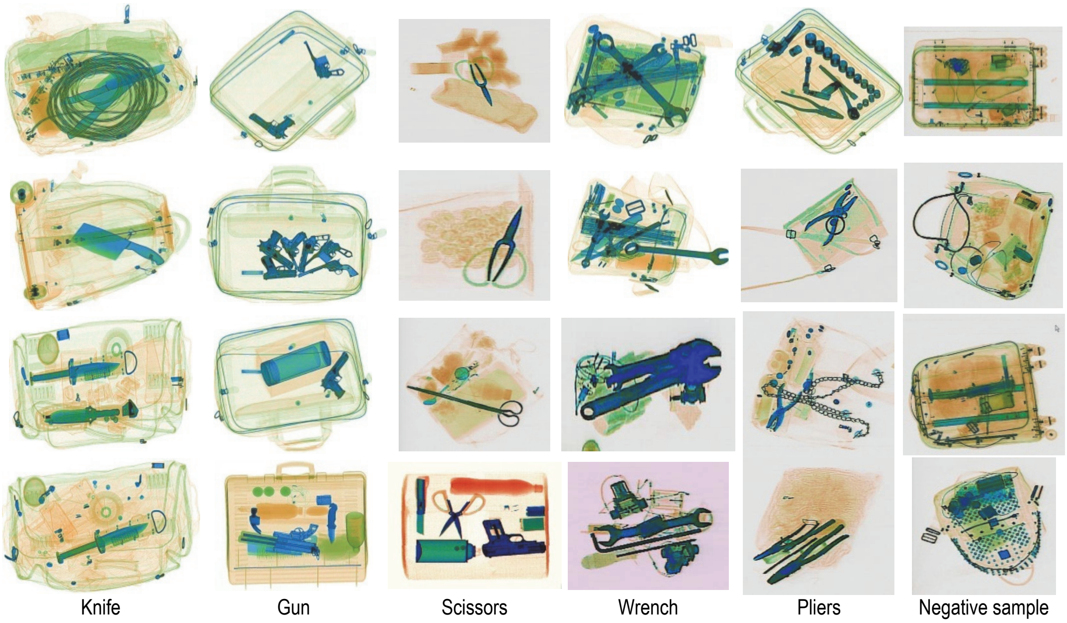


Figure 4 Some image examples in the subway security inspection X-ray scanning of data set SIXray

As for the evaluation indexes of detection performance, two indexes, detection Precision (P) and Recall (R) are adopted, and their definitions are shown in Eq. (2):

$$\begin{aligned} P &= \frac{TP}{TP+FP} \\ R &= \frac{TP}{TP+FN} \end{aligned} \tag{2}$$

In the above equation, TP represents the number of a certain type of dangerous goods that can be correctly detected as dangerous goods, FP represents the number of a certain type of non-dangerous goods that can be incorrectly detected as dangerous goods, and FN represents the number of a certain type of dangerous goods that can be incorrectly detected as non-dangerous goods. In general, Precision (P) and Recall *(*R) are mutually restrictive. In order to comprehensively evaluate detection performance, a PR curve with recall rate as abscissa and precision as ordinate is established, and the area under the PR curve is defined as Average Precision (AP). The larger the AP value, the better the performance of the model. In order to detect multiple objects, Average AP values of each class to obtain mean Average Precision (mAP). The larger the AP value is, the better the Average detection performance of the model for multiple objects. The above two indicators are defined as Eq. (3) and Eq. (4):

$$AP = \int_0^1 p(r)\,\mathrm{d}r \tag{3}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{4}$$

In terms of detection speed, Frames Per Second (FPS) is adopted as a speed evaluation index. The detection speed is closely related to the computing platform used. The detection experiment in this paper is conducted on a certain desktop computer. It has a 11th Gen Intel(R) Core (TM) i7 CPU @ 3.40 GHz, NVIDIA RTX 3090 GPU, and 24 GB GPU memory. The operating system is Ubuntu16.04, and CUDnn7.6.5 is used to support GPU acceleration. The Pytorch version 1.6.0 is adopted for deep learning framework, and the Python version 3.7 is used for programming.

4.2 Enhancement Processing of X-Ray Scan Image Dataset

Although the total number of images in SIXray's dataset is more than 1 million, the number of positive samples is only more than 8000, with a small amount of positive sample data and limited quality. In order to achieve better training performance, we preprocess the data in SIXray by mosaic data enhancement method. Firstly, we get the list of files in SIXray, and then four random integers no larger than the number of files are generated. These four random numbers are taken as the serial number of files, and corresponding pictures in the file list are read according to these serial numbers. The required data is generated by repeated image scaling, cutting, color gamut, and splicing process. It is particularly important to note that the relevant label data also needs to be processed, while processing the four images.

Images generated through the above process will own more rich background. It is equal to increase a lot of small objects in these images. The space semantic information of the images becomes rich. In addition, Four images combined is equivalent to increase the batch size, compared with the original data, reduced the hardware requirements of the model. And then the generalization performance of the model is improved. It should be noted that it is not necessary to process the data with many small objects by the above method, because more small objects will be added after processing, resulting in worse model generalization ability instead of improvement. The results of some images in SIXray dataset after data enhancement are shown in Fig. 5.

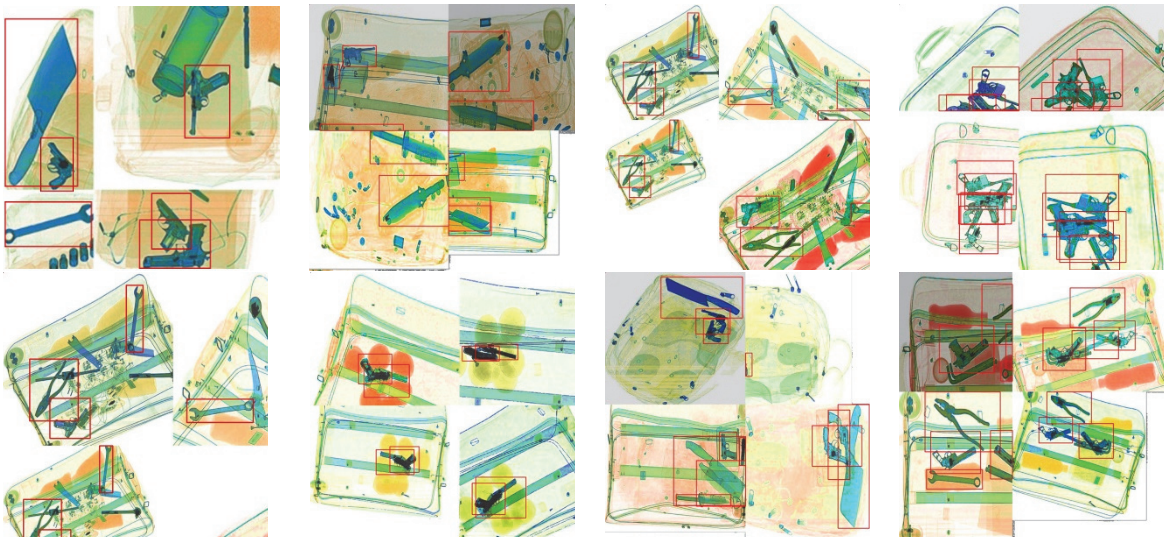


Figure 5 The result of data enhancement of some X-ray scanning images in SIXray dataset

Each image in Fig. 5 has its corresponding real anchor frame, which increases the number of real frame and enriches the scale of object after splicing. It can solve the problems of unbalanced data sets and is difficult to detect small objects, which is more conducive to the robustness of the training model.

4.3 Model Ablation Experiment and Results Analysis

Firstly, this section studies the influence of X-ray scanning image data set enhancement preprocessing on YOLOv3 model. The color image size of the dataset SIXray used in the experiment was 320 × 320 × 3, BatchSize was set to 64, and 400 epochs were trained in total. Adam optimizer is used for optimization. The experimental results of dataset enhancement are shown in Tab. 1.

As can be seen in Tab. 1, the preprocessing of data enhancement can greatly improve the detection precision of dangerous goods in baggage X-ray scan images, increasing mAP@0.5 value by 2.3%. Although the improvement in Precision (*P*) is only 0.5%, the improvement in Recall (*R*) is 3.42%. The results show that data enhancement preprocessing is suitable for this security inspection data set and can effectively alleviate the deficiency of less data sets and more small objects.

Table 1 Comparison of detection precision of different detection models before and after data enhancement preprocessing

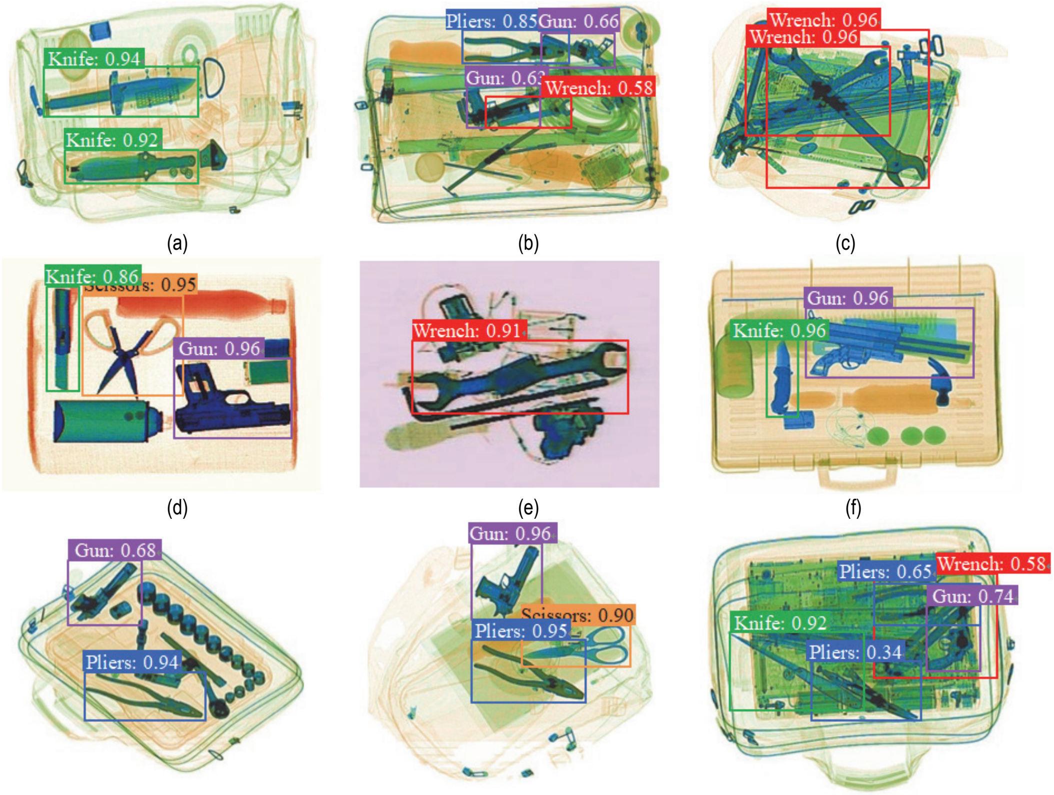
Detection model	mAP	Precision / %	Recall /%
YOLOv3	77.20	75.20	73.70
YOLOv3+ data enhancement	79.50	75.71	77.12

Therefore, the following experiments are all preprocessed with data enhancement before the model detection. In the following section, the influence of various improvement strategies for YOLOv3 model on detection precision is verified by the ablation experiments. Tab. 2 shows the ablation experiment settings and detection results.

Table 2 Comparison of experimental results of model ablation

YOLOv3	CCS-Darknet	SPP	FPN fusion	mAP@0.5 / %	FPS
√				77.20	281
√	√			83.11	326
√		√		79.92	265
√			√	82.70	318
√	√	√	√	88.62	392

As can be seen from Tab. 2, the mAP@0.5 value of the traditional YOLOv3 for the detection of dangerous goods in X-ray scanning images is 77.2%, and the detection speed is 281FPS. After the backbone network is improved to the cascaded cross-stage (CCS) Darknet, the value of mAP@0.5 reaches 83.11%, an increase of 5.91%, and the detection speed is also increased to 326, thus verifying that the cascaded cross-stage Darknet can enhance feature learning ability, reduce computing bottleneck and memory cost, and improve detection precision and speed. If the spatial pyramid pooling (SPP) module is added to the YOLOv3 model, mAP@0.5 can reach 79.92%, which increases by 2.72% compared with 77.2%. This indicates that the addition of SPP module can obtain different scale context information and increase the model receptive field to improve the feature system. In the aspect of feature fusion, after the FPN module is connected to a pyramid structure, mAP@0.5 is improved to 82.7%, owns an increase of 5.5%. At the same time, the detection speed was improved to 313FPS, 32 FPS higher than that of 281FPS, which once again verified the effectiveness of the cascade structure, and verified that FPN module cascade pyramid structure can integrate shallow spatial features in deep semantic features and enhance feature expression ability.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 6 Partial detection results of the CCS-YOLO v3 model on SIXray dataset

After the above three effective improvement strategies, the YOLOv3 model is improved to the new Cascade Cross Stage YOLO v3 model (CCS-YOLO v3) proposed in this paper. As can be seen from Tab. 2, the mAP@0.5 of this model reaches 88.62%, which is 11.4% higher than the traditional YOLOv3. Meanwhile, in terms of detection speed, the detection speed is improved to 392FPS, 111 FPS higher than the original YOLOv3's 281FPS. It can be concluded that the CCS-YOLOv3 model has a significant improvement in image detection speed and detection precision, and the model is effective and feasible. In order to demonstrate the detection performance of CCS-YOLOv3 model on dangerous goods in X-ray scanning images of security inspection, some detection results are selected and shown in Fig. 6. The numbers in the figure represent the confidence of the model to identify the object as a certain type of dangerous goods.

As can be seen from Fig. 6, the newly proposed CCS-YOLOv3 model can detect all dangerous goods in the 9 of X-ray security inspection images, without missing or wrong detection. In all 9 images, except Fig. 6e, the other images contain more than one type of dangerous goods. Even when two or more types of dangerous goods are present in the security inspection check images, the CCS-YOLOv3 test results are still good. Fig. 6c, e and i contain different styles of wrenches (wrenches vary in size and shape), even though wrenches can still be detected. The objects in Fig. 6b, c, e and i were overlapped and covered, and CCS-YOLOv3 could still correctly detect most of dangerous goods, even in the case of mutual interference of different objects. It is seen that the newly proposed CCS-YOLOv3 model achieves good detection effects on X-ray images in the above challenging real scenarios, and can frame all dangerous goods and predict the categories of dangerous goods with high confidence. It is worth pointing out that the detection precision of some dangerous goods will be seriously affected if there are large number of baggage and heavy overlapping coverage. For example, the pliers in Fig. 6i have a 0.34 probability of being identified as pliers.

4.4 Comparative Analysis of Detection Results

In order to compare the performance of CCS-YOLOv3 with other models more comprehensively, this model along with traditional YOLOv3, SSD and Faster-RCNN, is used for the detection of dangerous goods in X-ray scanning images, and the detection environment remains unchanged with the above experiments. SSD is a detection model based on feedforward convolutional neural network, and its network structure is divided into VGG16 feature extraction network and regression and classification sub-network. The obtained detection performance comparison is shown in Tab. 3.

Table 3 Comparison of detection performance of different models for dangerous goods in X-ray scanning images

Detection Model	mAP@0.5 / %	FPS	Model size / M
SSD	74.40	203	34.90
YOLOv3	77.20	281	36.20
Faster-RCNN	87.21	50	36.90
CCS-YOLO v3	88.62	392	36.60

As shown in Tab. 3, SSD has an obvious speed advantage compared with Faster RCNN, but its detection precision of 74.4% is lower than Faster RCNN's 88.62%. The detection precision of SSD and YOLOv3 is similar, and the detection speed of 203FPS is lower than that of YOLOv3 at 281FPS. However, the mAP@0.5 of Faster-RCNN can reach 87.21%, which is 10 percentage points higher than the traditional YOLOv3, but its detection speed of 50FPS is obviously inferior. Compared with the other three models, the CCS-YOLOv3 proposed in this paper has certain advantages in both detection precision and detection speed. In terms of model size, the current four models are comparable in size. It is worth pointing out that the model size of CCS-YOLOv3 model is adjustable, and the final size of the model can be determined by adjusting the number of X in CCS1_X or CCS2_X contained in the model. The detection precision will improve when the number of X increases, but the model size will also have corresponding increase.

5 CONCLUSIONS

A new cascade cross stage YOLOv3 model (CCS-YOLOv3) is proposed based on YOLOv3, and applied to the intelligent detection of security X-ray scanning images. Before applying the model to detection, the image data set is enhanced to improve the quality of the data set. Secondly, in order to improve the detection accuracy and speed, the backbone network of YOLOv3 is improved into a cascade cross stage Darknet network. After the backbone network, the spatial pyramid pooling module is added, and then the feature pyramid network module is connected in series with a bottom-up feature pyramid structure to realize feature fusion. The results of model Ablation experiments and baggage scanning image detection show that the average detection accuracy of CCS-YOLOv3 model is 11.4 percentage points higher than that of traditional YOLOv3, and the detection speed is 111 FPS higher, which verifies the effectiveness of the model improvement. Finally the CCS-YOLOv3 model is compared with the existing popular object detection models such as SSD and Faster-RCNN. The results show that the proposed model has certain advantages in detection accuracy and detection speed. The future research includes finding the most suitable model scale for image detection by adjusting the number of X in CCS1_X or CCS2_X contained in CCS-YOLOv3 model. In addition, the model's detection accuracy must be improved in the case of large number of baggage objects and serious overlapping coverage.

Acknowledgements

This study was supported by the natural science foundation of Gansu Province, China (20JR10RG307).

6 REFERENCES

[1] Michel, S., Hattenschwiler, N., Zeballos, M. et al. (2017). Comparing E-Learning and Blended Learning for Threat Detection in Airport Security X-Ray Screening. *2017 International Carnahan Conference on Security Technology (ICCST)*,1-6. https://doi.org/10.1109/CCST.2017.8167810

[2] Abidi, B., Page, D., & Abidi, M. A. (2005). A Combinational Approach to the Fusion, Denoising and Enhancement of Dual-Energy X-Ray Luggage Images. *2005 IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR), 112-120. https://doi.org/10.1109/CVPR.2005.386

[3] Xiaogang, Y. & Lirui, Y. (2012). A method on X-ray security image enhancement. *CT Theory and Applications*, *21*(4), 705-712.

[4] Bastan, M., Yousefi, M. R., & Breuel, T. M. (2011). Visual Words on Baggage X-Ray Images. *14th International Conference on Computer Analysis of Images and Patterns (CAIP)*, 360-368. https://doi.org/10.1007/978-3-642-23672-3_44

[5] Franzel, T., Schmidt, U., & Roth, S. (2012). Object Detection in Multi-View X-Ray Images. *Joint 34th DAGM (German Association for Pattern Recognition) and 36th OAGM Symposium (DAGM/OAGM)*, 144-154. https://doi.org/10.1007/978-3-642-32717-9_15

[6] Dalal, N. & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection.*2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 886-893. https://doi.org/10.1109/CVPR.2005.177

[7] Turcsany, D., Mouton, A., & Breckon, T. P. (2013). Improving Feature-Based Object Recognition for X-ray Baggage Security Screening Using Primed Visual Words. *2013 IEEE International Conference on Industrial Technology (ICIT)*, 1140-1145. https://doi.org/10.1109/ICIT.2013.6505833

[8] Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. *Computer Vision - ECCV 2006*, 404-417. https://doi.org/10.1007/11744023_32

[9] Hearst, M., Dumais, S., Osuna, E. et al. (1998). Support Vector Machines. *IEEE Intelligent Systems and their Applications*, *13*(4), 18-28. https://doi.org/10.1109/5254.708428

[10] Baştan, M. (2015). Multi-View Object Detection in Dual-Energy X-Ray Images. *Machine Vision and Applications*, 1045-1060. https://doi.org/10.1007/s00138-015-0706-x

[11] Mery, D., Riffo, V., Zuccar, I. et al. (2017). Object recognition in X-ray testing using an efficient search algorithm in multiple views. *Insight-Non-Destructive Testing and Condition Monitoring, 59*(2), 85-92. https://doi.org/10.1784/insi.2017.59.2.85

[12] Akcay, S. & Breckon, T. P. (2017). An evaluation of region based object detection strategies within X-ray baggage security imagery. *IEEE International Conference on Image Processing (ICIP)*, 1337-134. https://doi.org/10.1109/ICIP.2017.8296499

[13] Ren, S., He, K., Girshick, R. et al. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6),1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[14] Dai, J., Li, Y., He, K., & Sun, J. (2016). R-FCN: object detection via region-based fully convolutional networks. *30th International Conference on Neural Information Processing Systems (NIPS)*, 379-387. https://org/doi/10.5555/3157096.3157139

[15] Liang, K. J., Heilmann, G., Gregory, C. et al. (2018). Automatic threat recognition of prohibited items ataviation checkpoint with x-ray imaging: a deep learning approach. *Anomaly Detection and Imaging with X-Rays (ADIX) III*, 89-92. https://doi.org/10.1117/12.2309484

[16] Liu, W., Anguelov, D., Erhan, D. et al. (2016). SSD: Single shot multibox detector. *2016 European conference on computer vision*, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[17] Liang, K. J., Sigman, J. B., Spell, G. P. et al. (2019). Toward automatic threat recognition for airport X-ray baggage screening with deep convolutional object detection.

[18] Redmon, J., Divvala, S., Girshick, R. et al. (2016). You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 779-788. https://doi.org/10.1109/CVPR.2016.91

[19] Redmon, J. & Farhadi, A. (2018). Yolov3: An incremental improvement.

[20] Redmon, J. & Farhadi, A. (2017). YOLO9000: better, faster, stronger. *IEEE conference on computer vision and pattern recognition*, 6517-6525. https://doi.org/10.1109/CVPR.2017.690

[21] Sangwan, D. & Jain, D. K. (2019). An evaluation of deep learning based object detection strategies for threat object detection in baggage security imagery. *Pattern Recognition Letters*, *120*(4),112-119. https://doi.org/10.1016/j.patrec.2019.01.014

[22] See https://pjreddie.com/darknet/yolo/

[23]Akcay, S., Atapour-Abarghouei, A., & Breckon, T. P. (2018). GANomaly: Semi-supervised Anomaly Detection via Adversarial Training. *14th Asian Conference on Computer Vision*, 622-637. https://doi.org/ 10.1007/978-3-030-20893-6_39

[24] Akçay, S., Atapour-Abarghouei, A., & Breckon, T. P. (2019). Skip-GANomaly: Skip Connected and Adversarially Trained Encoder-Decoder Anomaly Detection. *2019 International Joint Conference on Neural Networks (IJCNN)*, 1-8. https://doi.org/10.1109/IJCNN.2019.8851808

[25] Gaus, Y. F. A., Akçay, N. S. et al. (2019). Evaluation of a Dual Convolutional Neu ral Network Architecture for Object-wise Anomaly Detection in Cluttered X-ray Security Imagery. *2019 International Joint Conference on Neural Network*, 37-52. https://doi.org/10.1109/IJCNN.2019.8851829

[26] 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. *2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2114-2123. https://doi.org/10.1109/CVPR.2019.00222

Contact information:

Jianjun WU, Associate Professor
(Corresponding author)
College of Information Technology and Communication, Hexi University,
No. 846, Huancheng North Road, Zhangye, Gansu Province, China
E-mail: wujj@hxu.edu.cn

Shaowen LIAO, Associate Professor
College of Information Technology and Communication, Hexi University,
No. 846, Huancheng North Road, Zhangye, Gansu Province, China
E-mail: wxdragon@hxu.edu.cn