Design of security image contraband detection system based on PP-YOLOE+_DCS

Fan Chen, Hua Jin*, Qinghan Li
Embedded System Laboratory,College of Engineering,Yanbian University,133002,Yanji,China

ABSTRACT

Traditional security screening methods mainly use manual identification of security images, there is a low identification of inefficiency, high error of judgement rate,which has become a bottleneck limiting public safety and security. Therefore, to deal with this problem, this paper proposes the security image contraband detection model PP-YOLOE+_DCS, which makes three main improvements on the basis of the PP-YOLOE+ model: To begin with, we introduced deformable convolution within the backbone network that strengthen the models' feature extraction capability; Secondly, we introduced a coordinated attention mechanism among the backbone network and the detection neck for better focusing the model on the object region;Finally, we replaced the original GIOU loss function with the SIOU loss function to improve the detection accuracy and training speed. The improved PP_YOLOE+_DCS model obtained achieves 91.4% detection accuracy,2.8% improvement compared with the baseline model mAP, only 0.24M additional parameters,and 420.2ms inference delay on embedded devices, which provides a new solution for the intelligence of contraband detection.

Keywords: X-ray security images, contraband detection, deformable convolution v2, coordinated attention

1. INTRODUCTION

As the demand for people's travel increases, the safety of public transportation hubs such as airports, train stations and subway stations is of great concern.X-ray baggage screening is one of the necessary measures to ensure people's travel security, which can detect and intercept people or items carrying contraband in a timely manner[1]. However, the traditional security screening method mainly relies on manual identification of contraband in X-ray images, which has low efficiency, time consuming, high false positive rate, and an accuracy rate of only 80-90% [2, 3]. Therefore, how to enhance the efficiency and precision of security checks has become a pressing issue in the field of security screening.

With the significant advances in image recognition and natural language processing made by deep learning techniques during recent years, many researchers and academics have begun to explore the application of deep learning to automatically detect prohibited items in X-ray security images.However, deep learning models usually require high-performance computers or cloud servers for training and inference, while embedded devices have unique advantages in deploying deep learning models by virtue of their small size, low power consumption, high performance and real-time response. Therefore, studying an X-ray security image contraband-assisted detection system that can be deployed in embedded devices to better assist security inspectors in their inspection work can effectively reduce the impact of human factors.

Considering the indicators of accuracy and model complexity, this paper selects the S-model model of PP-YOLOE+ as the baseline, and improves and optimizes it to obtain a deep learning model suitable for contraband detection. And using the Paddle Inference inference framework, the contraband detection model is deployed into the embedded device, aiming to realize the whole process of contraband detection in X-ray security images by intelligent means, reducing the work intensity of security inspectors and speeding up the efficiency of security inspection.

*812472694@qq.com

Seventh International Conference on Mechatronics and Intelligent Robotics (ICMIR 2023), edited by Srikanta Patnaik, Tao Shen, Proc. of SPIE Vol. 12779, 127791S · © 2023 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2688728

2. RELATED WORK

Recently, the advancement of deep learning techniques has enabled neural networks that can extract features directly from the initial data, which has significant advantages in terms of detection efficiency and accuracy. Object detection algorithms based on deep learning primarily fall into two-stage methods that are based on candidate regions and one-stage methods that are based on regression problems. The advantage of the two-stage method is high accuracy, and the disadvantage is high complexity and computational effort. The advantage of the one-stage method is its simplicity, speed and wide applicability, and the disadvantage is the relatively low detection accuracy.

Currently, X-ray image contraband detection methods are mainly based on targeted improvements and optimizations of the first-stage algorithms. While the improved model of contraband detection method brings some contribution to the contraband detection task, it also reveals some shortcomings. Zhou et al[4] proposed an improved X-ray safety image detection algorithm that is based on the YOLOv4 algorithm, which addresses the problems of complex backgrounds, target scale variations and mutual occlusion by introducing techniques such as deformable convolution, GHM loss and non-maximum suppression methods, with the disadvantage that GHM loss requires the computation of the gradient of the average value and the variance of the samples, thus increasing the computational sophistication and lengthening the training time.Ren et al[5] put forward the LightRay, a light-weight object detection framework for YOLOv4, that uses MobileNetV3 as its backbone feature extraction network, a light-weight feature pyramid network LFPN and CBAM attention mechanism, which is effective in achieving feature fusion at different scales and enhancing the feature information of small-sized contraband in complex backgrounds.However, when the detection results are unsatisfactory, more computational resources are still required.Liu et al[6] put forward a light-weight contraband detection method based on YOLOV4, denoted as LPD-YOLOv4. The method using MobilenetV3 as its backbone feature extraction network and using depth-separated convolution minimizes the parameter count and its computational consumption through optimizing the neck and head.In addition, it is designed with an self-adaptive space and channel attention blocks for improving the feature extraction capability.However, there is a reduced loss of detection accuracy compared to the original YOLOv4 model.Song et al[7] have proposed an enhanced YOLOv5 model that combines the CGhost module, Stem module,and Mixup data augmentation method to build contraband recognition capability. However, since Mixup requires the generation of new data samples,it increases the calculation sophistication for the model, which may contribute to the increase of training time and resources.

3. METHODOLOGIES

Because of the limited calculation performance of the deployed devices, the S-model of the PP-YOLOE+ series algorithm[8] is chosen as the baseline for this paper, on which optimization and improvement are carried out. The new model is denoted as PP-YOLOE+_DCS, and its algorithm structure is shown in Figure 1. The optimized PP-YOLOE+_DCS algorithm has three main improvement points, as shown in the solid box part in Figure 1, as follows:

(1) Introducing deformable convolution v2[9] that is to strengthen the backbone network's feature extraction capabilities and allow it to better capture the deformation and pose information of the target; (2)Introducing a CA (Coordination attention)[10] mechanism that enhances the attraction and perception of the model to key parts of the object among the backbone and feature fusion networks;(3)Substitute the former GIOU loss function of the model with the SIOU loss function[11], which is used to redefine the penalty index, and consider the fast convergence of the distance that the predicted box is from the real box to strengthen model's localization accuracy and matching ability of the object.

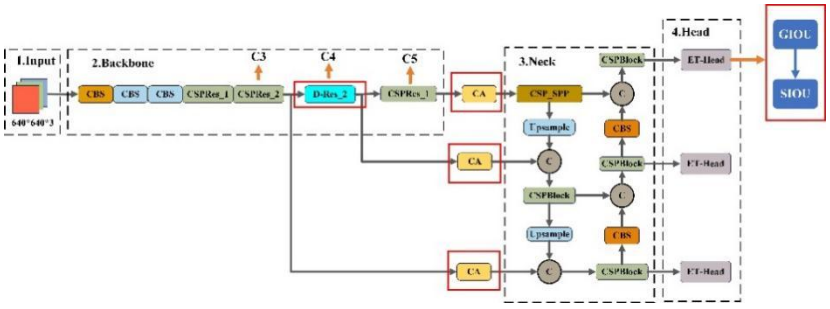

Figure 1. Structure of the improved PP-YOLOE+_DCS algorithm

3.1 Design of D-Res module

PP-YOLOE+ adopts an multi-scale feature map with three different scales to perform multi-scale feature integration which can adapt to targets of different sizes and have better detection effect on traditional natural images.However, due to the randomness of passenger luggage placement, after X-ray scanning, the contraband in the image will appear as multi-scale and multi-pose features, while ordinary items may be mistaken as contraband due to mutual occlusion and overlap between items.

To handle the above problem, we introduce the deformable convolutional DCNv2 that has modulation mechanism into the backbone network in this paper.Compared with the traditional convolution operation, DCNv2 can handle the spatial variation of target objects flexibly for improving the detection and recognition competence of the model. Since there is no fixed theoretical description for inserting deformable convolution in the network, the DBS module and the D-Res Block module are designed in this paper, and the D-Res module is obtained after replacing the backbone network CSPRes module, as shown in Figure 2. Each DBS module consists of DCNv2, BN and SiLU activation functions, for which n refers to the convolutional kernel size of DCNv2 and m refers to the sampling steps of the convolutional kernels.

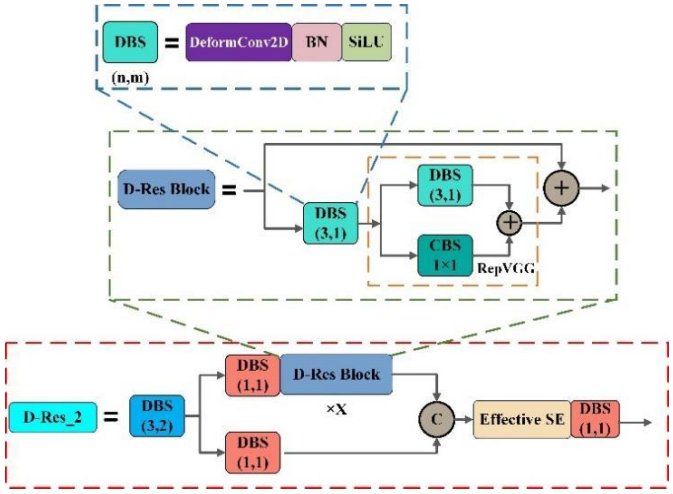

Figure 2. Design of D-Res module

3.2 Coordinate attention

Since security screening images are different from those in natural scenes, there is a large amount of noise and complex background information.This redundant information can hinder the overall network's representational capability and affect the model's overall detection of contraband.The coordinated attention mechanism is a method that can capture both channel relevance and spatial relevance in the feature map, and can adaptively adjust the weight relationship between different channels and different locations, so that the channels and locations with important information can receive more attention.It is shown that the coordination attention mechanism embedded into the backbone network and feature fusion network , shown on the CA module marked by the red solid line box in Figure 1.

The CA mechanism decomposes channel attention that captures features along two spatial directions from two one-dimensional feature encoding processes, one of which is responsible for capturing remote-dependent features and the other retains precise location information.The architecture is shown in Figure 3.

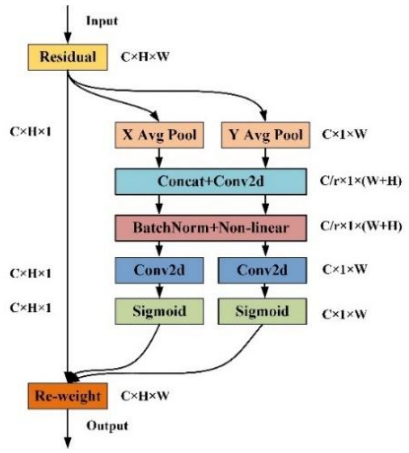

Figure 3. Architecture of the coordination of attention mechanism

The CA attention mechanism can be made up of two steps: Coordinated information embedding and Coordinated attention generation. The specific implementation is as follows:

①Coordinate message embedding

As for the input feature map $F \in R^{C\times H\times W}$, the average pooling operation using the pooling kernels of size $(H,1)$ and $(1,W)$ along the width and height directions respectively is performed so that the feature maps z_c^h and z_c^w in the width and height directions are obtained, which are calculated as shown in Equation (1).

$$z_c^h(h)=\frac{1}{W}\sum_{0\le i<W}x_c(h,i),\ z_c^w(w)=\frac{1}{H}\sum_{0\le j<H}x_c(j,w) \tag{1}$$

②Coordinate Attention generation

The feature maps which are encoded in the two directions of width and height are concatenated by the transformation of the information embedding step.Subsequently, the dimensionality is reduced to the original C/r by performing a 1×1 convolution operation in the channel dimension.After that, the batch normalized feature map will be fed into the Sigmoid activation function which results in a feature map of dimension $1\times(W+H)\times C/r$, as shown in Equation (2).

$$f=\sigma\left(F_1\left(\left[z^h,z^w\right]\right)\right) \tag{2}$$

After that, the feature map f is subjected to a 1×1 convolution operation following the initial width and height so that the feature maps f^h and f^w with the same dimension as the initial ones are obtained. Then, they are processed by the Sigmoid activation function to obtain the attention weights g^h and g^w on the width and height directions, respectively, as shown in Equation (3).

$$g^h=\sigma\left(F_h\left(f^h\right)\right),\ g^w=\sigma\left(F_w\left(f^w\right)\right) \tag{3}$$

Finally, the obtained attention weights are multiplied and weighted in accordance with the primal feature map so that a feature map y_c with attention weights is obtained, illustrated in Equation (4).

$$y_c(i,j)=x_c(i,j)\times g_c^h(i)\times g_c^w(j) \tag{4}$$

3.3 SIOU loss function

The PP-YOLOE+ model is calculated by adopting the GIOU loss function as the boundary box regression loss, although GIOU adds the minimum outer rectangle as the penalty term that is between the prediction box and the real box, and solves the problem that the gradient cannot be calculated when IOU is used as the loss function.However, there are still the following drawbacks: The first one is that when the prediction box and the real box overlap, the GIOU loss becomes

degraded to IOU loss, which cannot properly reflect the position of the prediction box within the real box and the height-to-width ratio of the prediction box. The second is that the gradient of the GIOU loss function also becomes very large when the intersection area of prediction box and real box is close to 0, which leads to unstable training.

Therefore, it is introduced that the SIOU loss function replaced the pre-existing GIOU loss function.The parameters associated with the SIOU loss function are shown in Figure 4.Where, B^{pred} denotes that the real box with the location of the center point $\left(b_{cx}, b_{cy}\right)$, B^{gt} denotes that the real box with the location of the center point $\left(b_{cx}^{gt}, b_{cy}^{gt}\right)$, σ is the distance between the center of the real box and the predicted box, w and h denoted that the width and height of the predicted box, w^{gt} and h^{gt} denoted that the width and height of the real box, respectively. The figure contains two rectangular boxes, one is the rectangle that is formed at the center of the real box and the prediction box, which is represented by the dashed box in Figure 4, and whose height and width differences are denoted as c_h and c_w, respectively; In the other one is the outer solid wireframe, which represents a minimum outer rectangle enclosing the real and predicted boxes, its height and width differences are denoted as ch' and cw', respectively.

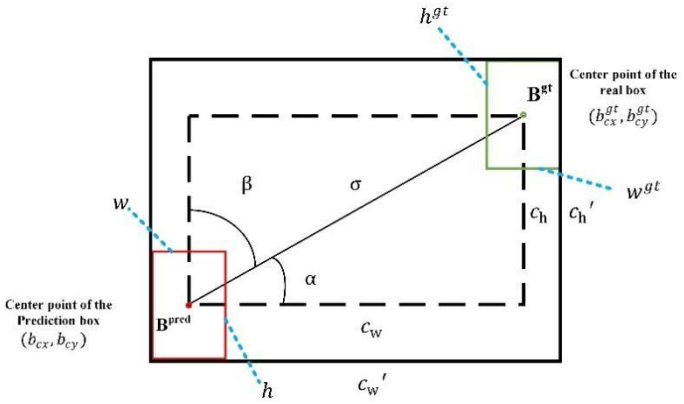

Figure 4. Schematic diagram showing the parameters of the prediction box and the real box

The SIOU loss function that is made up of four cost functions, and they are: angle cost, distance cost, shape cost and IoU cost.

(1) Angle cost. Its calculation formula are shown in (5).

$$\Lambda = 1 - 2 * sin^2(arcsin(x) - \frac{\pi}{4}) \tag{5}$$

Among them

$$x = \frac{c_h}{\sigma} = sin(\alpha) \tag{6}$$

$$c_h = max(b_{cy}^{gt}, b_{cy}) - min(b_{cy}^{gt}, b_{cy}) \tag{7}$$

$$\sigma = \sqrt{(b_{cx}^{gt} - b_{cx})^2 + (b_{cy}^{gt} - b_{cy})^2} \tag{8}$$

Where x is the sine function of angle α. The angle α can be obtained by taking the inverse function of the sine function. In the training process, if $\alpha < \frac{\pi}{2}$, then choose to minimize α, otherwise minimize β. Note that the angle loss will become 0 when the angle α is equal to $\frac{\pi}{2}$ or 0.

(2) Distance cost. In redefining the distance cost, the angle cost is to be considered, which is calculated by the formula shown in (9).

$$\Delta=\sum_{t=x,y}(1-e^{-\gamma\cdot\rho t}) \tag{9}$$

Among them

$$\rho_x=(\frac{b_{cx}^{gt}-b_{cx}}{c_w'})^2\ ,\ \rho_y=(\frac{b_{cy}^{gt}-b_{cy}}{c_h'})^2\ ,\ \gamma=2-\Lambda \tag{10}$$

From this equation, it can be seen that the effect of distance cost is greatly reduced when the angle α tends to 0 degrees, and becomes more pronounced when the angle α tends to 90 degrees.

(3) Shape Cost. Its calculation formula is shown in (11).

$$\Omega=\sum_{t=w.h}(1-e^{-\omega}t)^{\theta} \tag{11}$$

Among them

$$\omega_w=\frac{\left|w-w^{gt}\right|}{max\left(w,w^{gt}\right)}\ ,\ \omega_h=\frac{\left|h-h^{gt}\right|}{max(h,h^{gt})} \tag{12}$$

where the role of the θ parameter as a factor that controls the concern for shape cost, avoiding paying excessive attention to shape loss , and reducing the movement of the prediction box, its value range is $[2,6]$.

(4) IoU cost. Its calculation formula is shown in (13).

$$Loss_{IOU}=1-IoU \tag{13}$$

In conclusion, ultimately the SIOU loss function can be defined in Equation (14).

$$Loss_{SIOU}=1-IoU+\frac{\Delta+\Omega}{2} \tag{14}$$

4. EXPERIMENTS

4.1 Dataset source and production

For the training model, the dataset adopted is the OPIXray dataset[12] released by the Software Development Key Laboratory for Environment of Beijing University of Aeronautics and Astronautics in 2019.This dataset has a total of 8,885 X-ray images and 5 different knife categories. However, since the only contraband category labeled in this dataset is knives, the detection categories are single and limited, which is not suitable for the application.

Therefore, the following processing is taken to expand the dataset: in the first step, other contraband categories in this dataset are labeled using the labeling tool labelImg, including categories such as laptops, liquids, and cell phones; in the second step, considering the diversity of contraband categories in life, from the HiXray dataset[13] and the PIDray dataset[14], water, rechargeable batteries, cell phones , laptops, guns, lighters and aerosols. The new dataset, named OPI_MIXray, contains 12 categories of contraband and 30,676 images (training set:validation set:test set=8:1:1).

4.2 The experimental environment and hyperparameter settings

(1)Experimental environment: This paper uses AI Studio of Baidu Flying Pulp as the experimental platform, the hardware device has 32GB of video memory and memory capacity, 100GB of hard disk storage space, and the framework uses PaddlePaddle 2.4.0.

(2) Hyperparameter settings: The weight is updated via the adoption of a momentum optimizer and the initial momentum value that is 0.9. Learning rate that is set to 0.005 and dynamically adjusted using a cosine annealing learning rate decay and a linear preheat learning rate strategy with the first 5 rounds as preheat rounds.Batchsize is to be set to 32 and Epoch

is to be set to 300. After each 10 epochs in training, it is evaluated on the validation dataset and the optimal weights are saved. and save the optimal weights.

(3)Deployment test environment: Jeston Nano B01 suite with quad-core ARM A57 processor, 128-core MAXWELL GPU and 4GB LPD-DR memory, with sufficient AI computing power and model deployment framework of Paddle-Inference 2.4.1.

4.3 Model deployment

The model deployment framework uses Paddle-Inference inference framework. The framework is Baidu's open source high-performance inference engine based on the PaddlePaddle training model, which supports embedded devices such as x86 and ARM. The deployment process is shown in Figure 5. First, Paddle-Inference 2.4.1 is installed on the embedded device Jeston Nano, the trained and improved model is converted into a model deployment file and downloaded to Jeston Nano, and finally the initialization of the deployment file is completed through the API functions provided by the Paddle-Infernece framework, the model is invoked and inference is performed The inference result is obtained, and the location detection and classification of contraband is completed.

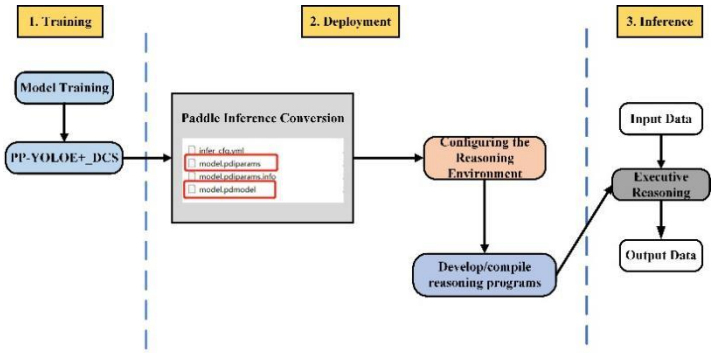

Figure 5. Model deployment inference flow chart

4.4 D-Res module introduces a comparison of location and attention mechanisms

In this paper, the CSPRes modules of C3, C4 and C5 stages of the backbone network output feature maps are replaced by D-Res modules, and the results under mAP@0.5 are all improved to different degrees than the baseline model, among which the C4 stage has the best results, it is shown in Table 1.

For the attention mechanism, we introduce ECA, CBAM and CA attention mechanisms respectively in this paper, which all have significantly improved the detection accuracy, among them the CA attention mechanism has the best effect, as shown in Table 1. Therefore, we choose to embed CA attention mechanism in the network in this paper, and replace the C4 stage CSPRes module with D-Res module at the output feature map of the backbone network, and adopt SIOU loss function to derive PP-YOLOE+_DCS model.

Table 1.D-Res module and attention mechanism experiments

D-Res module	**C3**	**C4**	**C5**	**Baseline**
mAP@0.5/%	90.2	**90.4**	89.8	88.6
Attentional Mechanisms	**CA**	**ECA**	**CBAM**	**Baseline**
mAP@0.5/%	**90.9**	89.7	90.4	88.6

4.5 Ablation experiments

The ablation of experiments in this paper as shown in Table 2. Where D represents the D-Res module, C represents the CA attention mechanism, and S represents the SIOU loss function. The experimental results demonstrate that all three improvement methods proposed in this paper can improve the detection precision to different degrees compared with the baseline model PP-YOLOE+_S. By introducing the D-Res module, it makes the convolutional kernel better adaptable to

contraband of different sizes, shapes and poses. By introducing the CA attention mechanism, it enables the network to acquire the context i southwest west of the region around the contraband in the security check image and reduce the influence of factors such as background and noise. By the replacement of the SIOU loss function, it reduces the localization error between the prediction box and the real box, which improves the detection accuracy and training speed of the model. Finally, with the combination of the three modules, mAP improves by 2.8% compared to the baseline model with only 0.24M additional number of parameters, and FLOPs and Latency both decrease to some extent.

Table 2. The ablation experiment of the improved model

Model	mAP@0.5/%	Params(M)	FLOPs(G)	Latency(ms)
Baseline	88.6	7.67	16.4	430.8
+D	90.4	7.80	15.6	410.7
+C	90.9	7.78	16.6	450.3
+S	90.3	7.67	16.4	432.2
+DC	90.7	7.91	15.8	419.3
+DCS	**91.4(+2.8)**	**7.91(+0.24)**	**15.8(-0.6)**	**420.2(-10.6)**

4.6 Comparison experiment

To further verify the improvement model PP-YOLOE+_DCS performance presented in this paper, the above model was selected for comparison experiments and also compared with other models of PP-YOLOE+ to better understand the relationship between model performance and scale. It can be seen from Table 3 that the improved method PP_YOLOE+_DCS introduced in this paper has reduced its inference delay by 10.6ms compared with that before the improvement, and good detection accuracy has been achieved in mAP@0.5:0.95 and mAP@0.5. This indicates that the improved method presented in this paper helps to improve the feature extraction ability of the model and makes it more capable of handling multi-scale and multi-attitude contraband. The improved model satisfies detection accuracy comparable to that of the complex model, making it more friendly for deployment on resource-limited devices and of practical application value.

Table 3. Results of experiments with different models

Model	OPI_MIXray			
	mAP@0.5/%	Params(M)	FLOPs(G)	Latency(ms)
Faster R-CNN	75.3	138.65	60.4	1586.6
SSD	62.3	2.55	2.9	76.2
YOLOv5_S	87.2	7.05	16.0	420.3
YOLOX_S	88.5	8.94	26.7	701.4
YOLOE+_S	88.6	7.67	16.4	430.8
YOLOE+_M	91.6	23.52	49.42	1298.1
YOLOE+_L	92.4	53.26	113.4	2978.8
YOLOE+_X	91.1	101.30	211.5	5556.7
Ours	**91.4(+2.8)**	**7.91(+0.24)**	**15.8(-0.6)**	**420.2(-10.6)**

4.7 Visualization of test results

Figure 6 shows the detection results for a partial test set (IoU=0.5). Among them, (a) and (c) show the detections results that are from the original PP-YOLOE+_S model, and (b) and (d) show the detections results that are from the improved PP-YOLOE+_DCS model. From the figures, it is observed that some targets with unclear shapes and contours are missed by the original model. And using the improved model, the missed targets can be detected (marked with red

boxes). Therefore, the improved model PP-YOLOE+_DCS submitted in it can better capture the feature information of contraband and improve the recognition and localization of contraband.

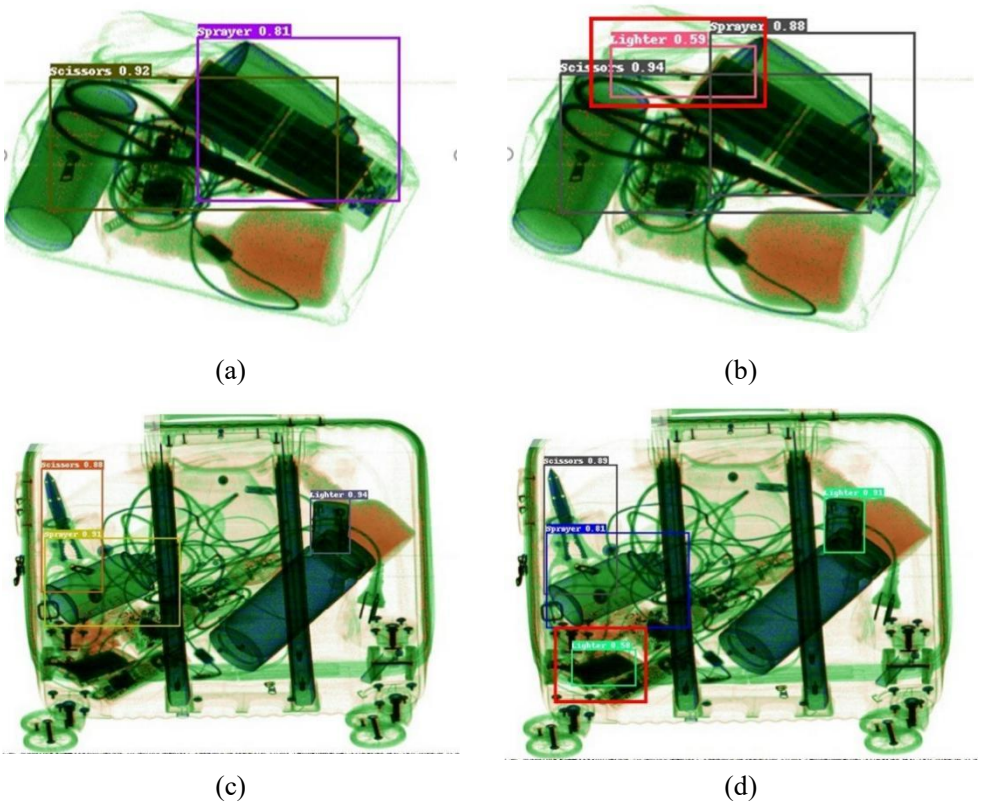

Figure 6. Comparison chart of test results

5. CONCLUSION

In order to better help security personnel detect contraband, we designed a method that is based on PP-YOLOE+_S for detecting prohibited items in X-ray security images in this paper. Targeted improvements are made to address the characteristics of prohibited items in the X-ray security screening images. DCNv2 is introduced in the backbone network and designed to obtain the D-Res module, which is used in C4 stage, so that it can enhance the accuracy of the model by better capturing the detailed information of prohibited items in X-ray security images with no increased computational effort. Moreover, we also introduce the CA attention mechanism of the backbone network and feature fusion network to further Improve the model's ability to locate prohibited items and extract critical characteristics. Finally, we replace the original of GIOU loss function with SIOU loss function to reduce the localization error and improve the detection accuracy. Through a series of tests, this paper designs a model that can improve the efficiency and accuracy of contraband recognition, which has engineering significance for the automation and intelligence of security image contraband detection.

ACKNOWLEDGMENTS

This research project was supported by the Education Department of Jilin Province ("Thirteen Five" Scientific Planning Project, Grant No.JJKH20180898KJ), Jilin Education Science Planning Project ("ThirteenFive" Plan, Grant No. GH170043), School-Enterprise Cooperation Program of Yanbian University. (No.YDXQ202301)

REFERENCES

[1] Akcay, S., Breckon, T., Towards automatic threat detection: A survey of advances of deep learning within X-ray security imaging, Pattern Recognition, 122, 108245 (2022).

[2] Wang, R., Shi, Y., Cai, M., Optimization and Research of Suspicious Object Detection Algorithm in X-ray Image. In 2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (Vol. 6, pp. 1357-1361). IEEE (2023).

[3] Ma, C., Zhuo, L., Li, J. et al., Occluded prohibited object detection in X-ray images with global Context-aware Multi-Scale feature Aggregation, Neurocomputing, 519, 1-16 (2023).

[4] Zhou, C., Xu, H., Yi, B., Yu, W., Zhao, C., X-ray security inspection image detection algorithm based on improved YOLOv4. In 2021 IEEE 3rd Eurasia Conference on IOT, Communication and Engineering (ECICE) (pp. 546-550). IEEE (2021).

[5] Ren, Y., Zhang, H., Sun, H., et al., LightRay: Lightweight network for prohibited items detection in X-ray images during security inspection, Computers and Electrical Engineering, 103, 108283 (2022).

[6] Liu, D., Liu, J., Yuan, P., et al., Lightweight prohibited item detection method based on YOLOV4 for x-ray security inspection, Applied Optics, 61(28), 8454-8461 (2022).

[7] Song, B., Li, R., Pan, X., Liu, X., Xu, Y.,. Improved YOLOv5 Detection Algorithm of Contraband in X-ray Security Inspection Image. In 2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI) (pp. 169-174). IEEE (2022).

[8] Xu, S., Wang, X., Lv, W., et al., PP-YOLOE: An evolved version of YOLO, arXiv preprint arXiv:2203.16250, (2022).

[9] Zhu, X., Hu, H., Lin, S., Dai, J., Deformable convnets v2: More deformable, better results. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9308-9316) (2019).

[10] Hou, Q., Zhou, D., Feng, J.,. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 13713-13722) (2021).

[11] Gevorgyan, Z., SIoU Loss: More Powerful Learning for Bounding Box Regression, arXiv preprint arXiv:2205.12740, (2022).

[12] Wei, Y., Tao, R., Wu, Z., Ma, Y., Zhang, L., Liu, X., Occluded prohibited items detection: An x-ray security inspection benchmark and de-occlusion attention module. In Proceedings of the 28th ACM International Conference on Multimedia (pp. 138-146) (2020).

[13] Tao, R., Wei, Y., Jiang, X., Li, H., Qin, H., Wang, J., et al., Towards real-world X-ray security inspection: A high-quality benchmark and lateral inhibition module for prohibited items detection. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10923-10932) (2021).

[14] Wang, B., Zhang, L., Wen, L., Liu, X., Wu, Y., Towards real-world prohibited item detection: A large-scale x-ray benchmark. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 5412-5421) (2021).