Engineering and Technology Journal e-ISSN: 2456-3358

Volume 09 Issue 08 August-2024, Page No.- 4805-4813

DOI: 10.47191/etj/v9i08.28, I.F. – 8.227

X-Ray Image Detection Algorithm for Prohibited Items Based on Feature Enhancement and Loss Optimization

ZHAO Xiaotao[1,2], LI Xinwei[1,2]

[1]School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China

[2]Henan Key Laboratory of Intelligent Detection and Control of Coal Mine Equipment, Jiaozuo 454000, China

ABSTRACT: To address the issue of low detection accuracy of prohibited items in X-ray security images caused by varying orientations, different scales, and the intertwining of targets with backgrounds, we propose a novel X-ray image detection algorithm based on feature enhancement and loss optimization. The model is built upon the ConvNext network and incorporates a Directional Channel Attention (DCA) mechanism, which efficiently captures the interaction information of local channels in different directions, thereby enhancing the accuracy of the detection model for prohibited items. Additionally, a multi-scale fusion bypass (MFB) branch is designed after the backbone network to integrate information from feature maps at different layers, thereby mitigating the interference of scale variations on the model. Furthermore, the loss function is redesigned to enable the model to automatically adjust its focus on hard samples, improving the overall detection performance. Experimental results on the SIXray dataset demonstrate that the proposed model achieves a mean Average Precision (mAP) of 91.43%, representing a 9.17% improvement over the original algorithm, thereby validating the effectiveness of the proposed method.

KEYWORDS: prohibited item detection; ConvNext; attention mechanism; multi-scale fusion; loss optimization

1. INTRODUCTION

Security screening is widely used in key public areas such as subways, trains, and airports, serving as a primary method to ensure public safety[1]. Currently, security checks primarily rely on manual inspection[2]. However, as public transportation systems continue to improve and the number of passengers increases annually, especially during peak travel periods, the complexity of items in luggage may lead to errors and omissions by security personnel[3]. Therefore, achieving accurate and rapid automatic identification of prohibited items has become critically important in security screening[4].

With the rapid development of deep learning in object detection, convolutional neural networks (CNNs) have been actively applied to X-ray prohibited item recognition[5]. For example, Miao et al.[6] proposed the ResNet-CHR model based on the ResNet[7] network, designing three branches to extract features at different scales, thereby enhancing detection performance under severe occlusion conditions. Wei et al.[8] introduced a De-Occlusion Attention Module (DOAM), inspired by the varying imaging colors of X-rays across different materials; this module focuses on the edge and material information of prohibited items, generating high-quality feature maps for detection. Shaoqing Yao et al.[9] designed asymmetric convolution and dilated convolution modules in both deep and shallow feature layers to reduce misclassification and improve segmentation accuracy. Xia Feng et al.[10] proposed an X-ray security image classification method to address sample imbalance, using the ResNet101 residual network as a baseline and introducing a cost-sensitive factor into the loss function, effectively mitigating the issue of sample imbalance for prohibited items. Yang Cao et al.[11] tackled the challenges of scale variation and occlusion in prohibited item detection by incorporating deformable convolutions and attention mechanisms, thereby improving the model's detection accuracy. Jinhao Yuan et al.[12] employed large-kernel attention to extract low-level features of images and introduced the CBAM attention module into the backbone network to enhance information in regions of interest, further boosting the model's detection accuracy. While these algorithms have improved the detection performance of prohibited items to some extent, challenges remain in real-world security scenarios due to the complex backgrounds, significant scale variations, and uneven distribution of feature information in X-ray security images, which continue to limit the accuracy of these methods.

To address the aforementioned challenges, this paper optimizes the ConvNext[13] network as the baseline. The main contributions are as follows:

a) Design of the Directional Channel Attention (DCA) Mechanism: This attention mechanism is designed to enhance the network's focus on channels in different directions within prohibited item images, effectively suppressing irrelevant information and extracting the underlying features of prohibited items.

b) Introduction of a Multi-Scale Fusion Bypass Branch (MFB): By incorporating this bypass branch, the model can fully leverage the semantic information

from different feature layers, thereby improving its adaptability to variations in object scale.

c) Development of the GF Loss Function: Based on the Focal Loss[14] function, the GF Loss function increases the weight of hard samples in the loss calculation while dynamically adjusting the model's attention to these difficult samples.

2. MODEL OPTIMIZATION

To address the issue of low detection accuracy for prohibited items, this paper uses the ConvNext model as the baseline. First, a Directional Channel Attention (DCA) mechanism is added at the end of each ConvNext Block in the final stage of the backbone, enhancing the model's ability to perceive critical features of prohibited items. Additionally, a Multi-Scale Fusion Bypass (MFB) branch is introduced to improve the model's sensitivity to prohibited items of varying scales. Finally, the Cross-Entropy (CE) loss function is replaced with the GF Loss function to enhance the model's convergence speed and detection accuracy. The structure of the optimized ConvNext network is illustrated in Figure 1

.

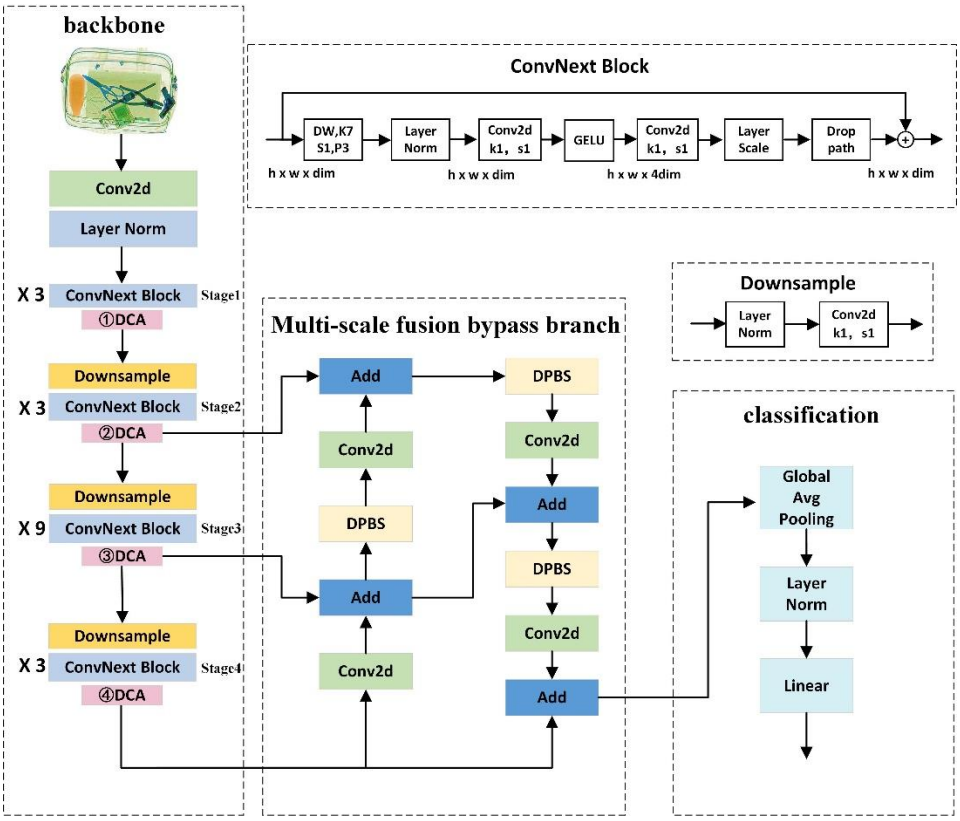

Fig.1 Optimized ConvNext structure diagram

2.1 Directional Channel Attention Mechanism

In the field of computer vision, attention mechanisms learn weight distributions that allow networks to dynamically adjust the importance of different feature channels, aiding in the extraction of relevant information. Hu et al.[15] introduced the Squeeze-and-Excitation Networks (SE-Net), which applies two-dimensional global average pooling to aggregate features and then performs adaptive dimensionality reduction to capture channel dependencies. However, SE-Net does not fully consider the information between channels in different directions. Subsequently, Hou et al.[16] proposed the Coordinate Attention Network (CA-Net), which incorporates directional information into the channel attention framework. CA-Net uses two-dimensional convolutions to concatenate and then separate channels in horizontal and vertical directions. While this approach reorganizes channels, it does not adequately capture dependencies between local channels across different directions.

To comprehensively capture the interaction and dependencies between channels in various directions for prohibited items, this paper designs the Directional Channel Attention (DCA) mechanism. The structure of the DCA mechanism is illustrated in Figure 2

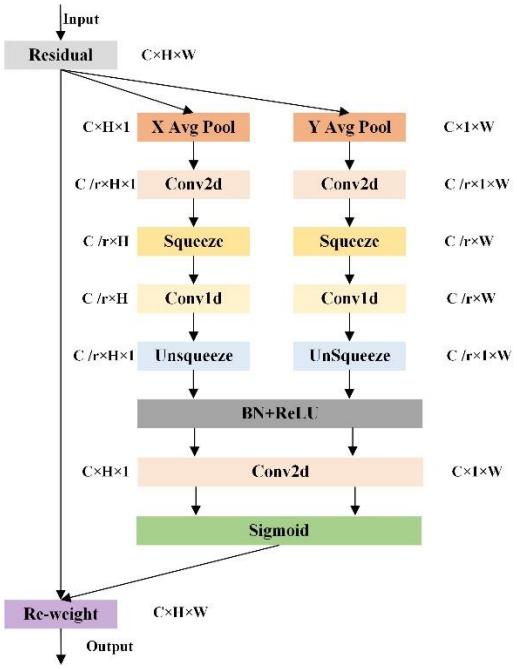

Fig.2 DCA attention mechanism

Traditional channel attention mechanisms use two-dimensional global average pooling to obtain channel features, which may lead to the neglect of some important information. In contrast, the Directional Channel Attention (DCA) mechanism performs global average pooling along both the width and height directions of the input feature map $D \in R^{C\times H\times W}$, extracting aggregated features in each direction for the prohibited item images. The feature expression is given as follows:

$$D_c^h(h) = \frac{1}{W}\sum_{j=1}^{w} u_c(h, j) \quad (1)$$

$$D_C^w(w) = \frac{1}{H}\sum_{i=1}^{H} u_c(i, w) \quad (2)$$

In Equation (1), $D_c^h(h)$ represents the output of the c channel with a height of h, while W denotes the width of the input U.

In Equation (2), $D_c^w(w)$ represents the output of the c channel with a width of w, and H denotes the height of the input U.

Next, to reduce the number of parameters in the model, a 1×1 convolution function f_1 is used to decrease the number of channels to C/r of the original amount. Subsequently, dimensionality reduction is applied along both the width and height, resulting in matrices $D(h) \in R^{C/r\times H}$ and $D(w) \in R^{C/r\times W}$ that capture channel information in different directions, as shown in Equations (3) and (4).

$$D(h) = F_{s1}(f_1(D^h)) \quad (3)$$

$$D(w) = F_{s2}(f_1(D^w)) \quad (4)$$

In the equations, F_{s1} and F_{s2} represent the results of dimensionality reduction along the width and height, respectively.

To efficiently capture the feature dependencies between channel dimensions in different directions and facilitate local cross-channel information interaction, this paper employs one-dimensional convolution to obtain channel weights for prohibited item image features along the height and width. Subsequently, an expansion operation is performed to restore the dimensions. The specific computations are detailed in Equations (5) and (6).

$$B(h) = F_{e1}(CD^{K1}(D(h))) \quad (5)$$

$$B(w) = F_{e2}(CD^{K2}(D(w))) \quad (6)$$

In the equations, F_{e1} and F_{e2} represent the expansion operations applied along the width and height, respectively. CD denotes the one-dimensional convolution, while $K1$ and $K2$ are the corresponding adaptive convolution kernels.

Simultaneously, F_{e1} and F_{e2} undergo batch normalization and ReLU activation to accelerate network convergence and enhance the model's generalization capability. Following this, a 1×1 convolution function f_2 is applied to restore the number of channels. After applying the sigmoid activation function, the channel attention weights for the feature map H are obtained as $P(h) \in R^{C\times H\times 1}$, and for the width dimension W are obtained as $P(w) \in R^{C\times 1\times W}$. The expressions for these weights are given in Equations (7) and (8).

$$P(h) = \sigma(\varphi(bn(B(h)))) \quad (7)$$

$$P(w) = \sigma(\varphi(bn(B(w)))) \quad (8)$$

In the equations, bn represents batch normalization, φ denotes the ReLU activation function, and σ refers to the sigmoid activation function.

Finally, the feature maps $P(h) \in R^{C\times H\times 1}$, $P(w) \in R^{C\times 1\times W}$, and $D \in R^{C\times H\times W}$ are multiplied element-wise with the input feature map to capture semantic information in different directions and assign distinct weights to each feature channel. The final result is given by Equation (9).

$$D_{out} = D \otimes P(h) \otimes P(w) \quad (9)$$

The DCA attention mechanism increases the channel weights associated with prohibited item features, further enhancing the model's ability to extract features from images with complex backgrounds.

2.2 Multi-scale Fusion Bypass Branch

In security inspection images, the placement of prohibited items often results in significant variations in object scale. The same type of prohibited item may appear in different sizes and styles within the images. The ConvNext backbone network has certain limitations when handling scenarios with such scale variability, primarily due to the repeated stacking of blocks and downsampling operations. While stacking multiple blocks can enhance the model's feature extraction capability, it also restricts the receptive field of neurons, making it difficult for the network to effectively capture contextual information. Additionally, downsampling reduces the size of the feature maps, which in turn diminishes the network's ability to process fine details when learning and recognizing features.

To improve the model's ability to recognize objects with varying scales, this paper introduces a Multi-scale Fusion Bypass Branch (MFB) to fully leverage the information across different feature layers. As illustrated in Figure 3

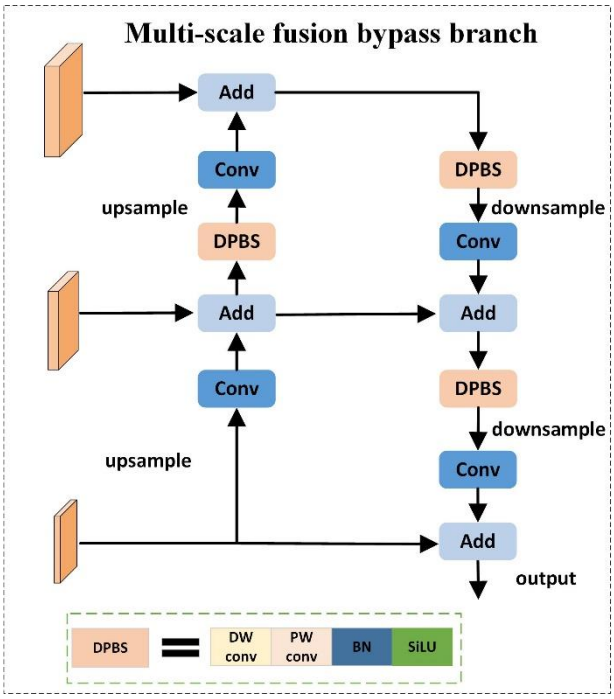

Fig.3 Multi-scale fusion bypass branch

First, the output of the ConvNext backbone network is upsampled using bilinear interpolation, followed by a 1×1 convolution to match the output of the feature layers.

Subsequently, adjacent feature layers are fused using an addition (add) operation rather than concatenation (concat), which increases the feature information without altering the

Dimensionality, thereby saving computational resources. However, since some information is shared between the high- and low-level feature maps, additive fusion can lead to redundant information. To address this issue, a DPBS module, comprising depthwise convolution, pointwise convolution, batch normalization, and the SILU activation function, is introduced. In pointwise convolution, each kernel interacts only with a single channel without influencing others, which helps to reduce feature overlap, while depthwise convolution lowers the resolution, preserving essential feature information. The inclusion of the DPBS module not only eliminates the overlapping effect resulting from feature fusion but also enhances the model's nonlinear representation capabilities.

The global features from the higher layers are propagated to the lower layers through upsampling and combined with local features, resulting in a fused feature map that contains multi-scale information. By merging local and global features, the MFB effectively captures a more comprehensive range of scale variations in prohibited items, thereby mitigating the impact of scale differences on the model's detection accuracy.

2.3 Loss Function Improvement

The Focal Loss (FL) function, compared to the traditional Cross Entropy (CE) loss function, introduces a modulation factor γ to increase the weight of hard-to-classify samples in the loss function, thereby improving the model's detection performance on these difficult samples, as shown in Equations (10) and (11).

$$FL = -\alpha_t (1-p_t)^{\gamma} \log(p_t) \qquad (10)$$

.

$$p_t = \begin{cases} p & y=1 \\ 1-p & other \end{cases} \qquad (11)$$

In these equations, α_t represents the balancing factor, γ is the modulation factor, p denotes the predicted value, and y is the true label.

Figure 4 illustrates the changes in average precision (AP) for five types of prohibited items over 300 iterations when using the FL function with the baseline ConvNext network

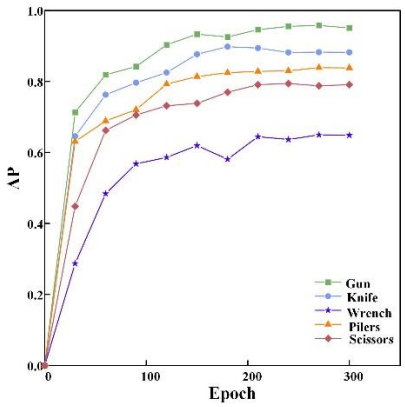

Fig.4 Contraband detection accuracy change curve

As observed in Figure 4, the overall AP value for the wrench is relatively low. However, as the number of iterations increases, the AP value for the wrench gradually stabilizes, indicating that the model's ability to extract features from this difficult sample has improved. At this point, maintaining a constant γ could cause the model to overly focus on the originally difficult samples, reducing its attention to other samples and thereby affecting overall detection accuracy. To address this issue, we propose the GF Loss function, which builds upon the Focal Loss function, as shown in Equations (12) and (13).

$$GF = -\alpha_t (1-p_t)^{\beta} \log(p_t) \qquad (12)$$

$$\beta = \gamma * e^{-\frac{epoch}{epochs}} \qquad (13)$$

In these equations, *epoch* represents the current training epoch, and *epochs* denotes the total number of training epochs.

By designing a novel modulation factor β, the impact of previously difficult samples on the loss can be dynamically reduced, thereby enhancing the model's ability to comprehensively recognize prohibited items.

3. EXPERIMENTIAL RESULTS AND ANALYSIS

3.1 Experimental environment and dataset

The experiments were conducted on an Ubuntu 20.04 system, using Python 3.8. The hardware utilized includes an Intel Core i7-8700K @ 3.7 GHz six-core CPU, 16 GB of RAM, and an NVIDIA GeForce RTX 2080 GPU with 8 GB of VRAM. The batch size was set to 8, and the AdamW optimizer was used with an initial learning rate of 5×10^{-4}. The number of epochs was set to 300.

The experiments utilized the SIXray and OPIXray datasets for validation. The SIXray dataset comprises 8,929 images of prohibited items, including guns, knives, wrenches, pliers, and scissors, with sample images shown in Figure 5. To enhance data diversity, data augmentation techniques such as random cropping, rotation, flipping, translation, and brightness adjustment were applied to these five categories of prohibited items. Finally, the augmented images were randomly divided into training and testing sets in an 8:2 ratio.

The OPIXray dataset, generated using software, contains 8,885 X-ray images of prohibited items, including five types of knives: folding knives, straight knives, scissors, utility knives, and multi-tools. In the experiments, the dataset was divided into training and testing sets, with the training set comprising 80% of the images (7,109 images) and the testing set comprising 20% (1,776 images).

All experiments were conducted using the SIXray dataset, with the SIXray and OPIXray datasets being used together only for algorithm comparison experiments.

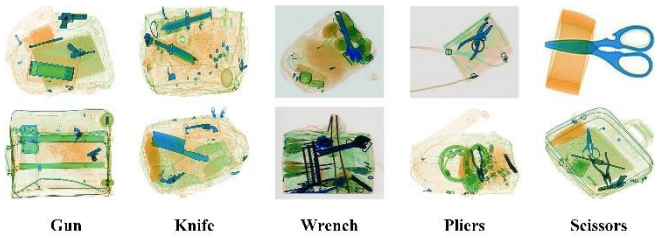

Fig.5 Data sample example

3.2 Evaluation Metrics

This study uses the following evaluation metrics to comprehensively assess model performance: Detection Precision (AP), Mean Average Precision (mAP), the number of Parameters, and Floating Point Operations (FLOPs). The definitions of AP and mAP are given in Equations (14) and (15).

$$AP = \frac{TP}{TP + FP} \tag{14}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{15}$$

In the equations, TP represents the number of true positive samples, where the actual class is positive and the prediction is also positive; FP denotes the number of false positive samples, where the actual class is negative but the prediction is positive; AP_i is the average precision for the i class of prohibited items; and N denotes the total number of prohibited item classes.

3.3 Comparison of Attention Mechanisms

To investigate the impact of different attention mechanisms on the network's ability to identify prohibited items, this experiment incorporated various attention mechanisms at the end of the ConvNext Block residual connections and compared them against the baseline ConvNext model. The detailed results are presented in Table 1.

Tab.1 Comparative test of attention mechanism

Attention	Params/M	FLOPs/M	mAP/%
-	28.57	4455.53	82.26
SE	28.67	4456.20	84.70
CBAM	28.76	4456.90	84.00
CA	28.71	4457.11	85.11
DCA(ours)	28.68	4456.77	86.78

From the results presented in Table 1, it can be observed that incorporating the DCA attention mechanism led to a 4.52% improvement in mAP with only a minimal increase in the number of parameters and computational cost. Compared to SE, CBAM, and CA, DCA achieved the most significant enhancement in mAP, demonstrating its effectiveness.

3.4 Comparison of Loss Functions

To validate the effectiveness of the GF loss function, the experiment compared CE, FL, and GF loss functions using the ConvNext algorithm as the baseline.

Tab.2 Loss function comparison experiment

Loss function	AP(%)					mAP(%)
	Gun	Knife	Wrench	Pliers	Scissors	
CE	95.84	88.62	63.89	83.57	79.42	82.26
FL	95.67	88.12	70.37	85.72	80.71	84.11
GF	96.43	91.18	69.21	85.96	82.89	85.13

According to the results in Table 2, the model using the GF loss function achieved a 2.87% and 1.02% higher mAP compared to the CE and FL loss functions, respectively. This demonstrates the effectiveness of the dynamic modulation factor in enhancing the accuracy of the loss function. Figure 6 shows the loss curves for the three loss functions under the same experimental conditions.

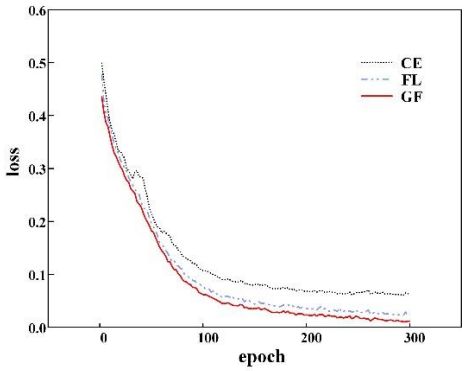

Fig.6 Loss decline curve

From Figure 6, it can be observed that the model using the GF loss function demonstrates superior performance in both convergence speed and loss value compared to the CE and FL loss functions, further validating the advantages of the GF loss function.

3.5 Algorithm Comparison Experiment

To evaluate the performance of the proposed algorithm in prohibited item detection, this study compares it with several state-of-the-art algorithms.

To ensure fairness in the experiments, all algorithms were trained under the same experimental conditions, using the same datasets and data augmentation techniques as those employed for the proposed algorithm. Detailed comparison results are presented in Tables 3 and 4

.

Tab.3 SIXray dataset comparison experiment

	AP/%					mAP/%
	Gun	Knife	Wrench	Pliers	Scissors	
CHR+ResNet101	97.56	88.73	75.26	90.40	81.77	86.74
Swin Transformer	93.38	89.23	65.31	86.06	79.40	82.67
DenseNet121	92.99	86.98	70.01	88.46	83.75	84.43
Shufflenet V2	93.52	87.63	81.63	89.29	83.39	83.78
MobileNet V3	96.04	87.96	64.92	86.64	80.19	83.15
EfficientNet V2	97.07	87.41	70.81	84.74	81.01	84.20
ConvNext	95.84	88.62	63.89	83.57	79.42	82.26
ConvNext-M(ours)	98.07	94.61	82.11	93.44	88.94	91.43

Tab.4 OPIXray dataset comparison experiment

	AP/%					mAP/%
	Folding knife	Straight knife	Scissors	Utility knife	Multi -tool knife	
CHR+ResNet101	58.13	86.57	77.33	60.24	85.02	73.46
Swin Transformer	67.14	79.67	71.42	62.35	70.30	70.18
DenseNet121	66.25	80.63	84.02	55.44	60.35	69.34
Shufflenet V2	64.32	75.00	86.21	48.22	88.33	72.41
MobileNet V3	71.08	86.20	76.55	56.74	72.57	72.63
EfficientNet V2	60.43	77.04	72.51	50.36	72.21	66.51
ConvNext	72.65	78.37	76.68	62.18	71.24	72.26
ConvNext-M(ours)	72.42	78.36	80.23	68.29	76.17	75.21

As shown in Table 3, the proposed ConvNext-M algorithm significantly improves mAP by 9.17% compared to the original ConvNext network. Notably, the AP for wrenches and pliers increased substantially by 18.22% and 9.87%, respectively, indicating enhanced capability in extracting key features of prohibited items. Compared to other algorithms, ConvNext-M exhibits superior mAP and performs exceptionally well in terms of average precision (AP) across all five categories of prohibited items, confirming the advantages of the proposed algorithm in prohibited item detection.

Analysis of Table 4 reveals that the ConvNext-M algorithm demonstrates the best detection performance on the OPIXray dataset compared to other algorithms. This result

confirms the robust generalization capability of the proposed algorithm.

3.6 Ablation experiments

To validate the effectiveness of the three proposed optimizations for improving prohibited item detection, we conducted an ablation study using the ConvNext algorithm as the baseline. The optimizations were introduced incrementally, and the specific experimental results are detailed in Table 5.

Tab.5 Ablation experiment

DCA	MFB	GF Loss	Params/M	mAP/%
-	-	-	28.57	82.26
✓	-	-	28.68	86.78
	✓	-	30.25	86.20
✓	✓	✓	30.36	91.43

As shown in Table 5, the introduction of the DCA attention mechanism alone resulted in a 4.52% increase in mAP, with nearly no change in the number of parameters. When the MFB structure was added individually, the mAP improved by 3.94%, with a slight increase in parameters. After incorporating all three proposed optimizations, the model achieved a 9.17% increase in mAP with only a 1.79M increase in parameter count. This demonstrates that the proposed optimizations significantly enhance detection accuracy with a relatively modest increase in parameter complexity, validating their effectiveness in improving prohibited item detection performance.

3.7 Prohibited Item Localization

To validate the network's performance in prohibited item localization, this study employs the Grad-CAM visualization technique. By using a weakly supervised approach, heatmaps are overlaid on the original images to illustrate the network's localization capabilities. Compared to traditional bounding box annotations, this method provides a more intuitive and convenient way to demonstrate the network's localization performance. As shown in Figure 7, the ConvNext-M network, with its effective feature extraction and fusion strategies, exhibits superior localization results compared to the ConvNext network.

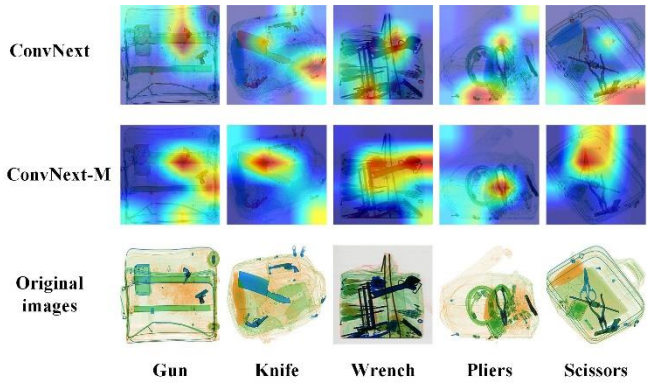

Fig.7 Contraband location effect

4. CONCLUSIONS

This study builds upon the ConvNext model to propose an X-ray image prohibited item detection algorithm based on feature enhancement and loss optimization. By designing a Directional Channel Attention (DCA) mechanism, introducing a Multi-scale Fusion Bypass Branch (MFB), and optimizing the loss function, the model's detection capabilities for various prohibited items have been significantly improved. Extensive experiments demonstrate that the optimized model shows a marked enhancement in detection performance compared to the original model, effectively meeting the demands of real-world security screening scenarios. Future work will focus on achieving model lightweighting while maintaining performance, to further enhance the efficiency of the algorithm.

REFERENCES

1. MU Siqi， LIN Jinjian， WANG Haiquan， WEI Xiongzhi. An Algorithm for Detection of Prohibited Items in X-ray Images Based on Improved YOLOv4[J]. Acta Armamentarii, 2021, 42(12): 2675-2683. (in Chinese).
2. Cheng Lang, Jing Chao, Chen Wenpeng. LLP-NAS: Prohibited Item Detection Algorithm with Neural Network Architecture Search using X-ray Images[J]. Science Technology and Engineering,2024,24(2):665-675. (in Chinese).
3. Wei, Yanlu, et al. "Occluded prohibited items detection: An x-ray security inspection benchmark andde-occlusion attention module." Proceedings of the28th ACM international conference on multimedia. 2020.

4. Ke Zhang, Liang Zhang. Multi-Scale Detection for X-Ray Prohibited Items in Complex Background[J]. Laser & Optoelectronics Progress, 2021, 58(22): 2210002. (in Chinese).
5. Akcay S , Kundegorski M E , Willcocks C G ,et al.Using Deep Convolutional Neural Network Architectures for Object Classification and Detection Within X-Ray Baggage Security Imagery[J].IEEE Transactions on Information Forensics and Security, 2018:2203-2215.DOI:10.1109/TIFS.2018.2812196.
6. Miao C, Xie L, Wan F, et al. Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.2019: 2119-2128.
7. He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
8. Wei, Yanlu, et al. "Occluded prohibited items detection: An x-ray security inspection benchmark andde-occlusion attention module." Proceedings of the28th ACM international conference on multimedia. 2020.
9. Shaoqing Yao, Zhigang Su. Prohibited Item Identification Algorithm Based onLightweight Segmentation Network[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0210022. (in Chinese).
10. FENG X，WEI X K，LIU C H，et al. Contraband classification method for X-ray security images considering sample imbalance[J]. Journal of Beijing University of Aeronautics and Astronautics，2023，49（12）：3215-3221（in Chinese） doi: 10.13700/j.bh.1001-5965.2022.0095
11. Yang Cao, Li Zhang, Junxi Meng, Qian Song, Letian Zhang. Multi-Target Prohibited Item Recognition Algorithm for X-Ray Security Scene[J]. Laser & Optoelectronics Progress, 2022, 59(10): 1015009.(in Chinese).
12. YUAN Jinhao, ZHANG Nanfeng, RUAN Jieshan, GAO Xiangdong. Detection of prohibited items in X-ray images based on modified YOLOX algorithm[J]. Laser Technology, 2023, 47(4): 547. (in Chinese).
13. Liu, Zhuang, et al. "A convnet for the 2020s." Proceedings of the IEEE/CVF conference on computer vision and pattern recogntion. 2022.
14. Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.
15. Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
16. Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.