DOI: 10.19884/j.1672-5220.202501013

WaveMFFM: Wavelet-Guided Multi-Feature Fusion Module for X-Ray Prohibited Item Detection

SUN Peng, CHEN Guangfeng*
College of Mechanical Engineering, Donghua University, Shanghai 201620, China

Abstract: To improve the accuracy of detecting prohibited items in X-ray images, this study proposes a wavelet-guided multi-feature fusion module (WaveMFFM), an easy-to-integrate, plug-and-play module that can be seamlessly incorporated into existing detectors. WaveMFFM innovatively introduces the wavelet transform and pioneers the de-occlusion wavelet convolution (DOWC) structure, which dynamically integrates low-frequency global contour information and high-frequency detailed texture features through a frequency-domain decoupling mechanism. This approach effectively resolves the feature confusion issue inherent in conventional convolutional operations under occlusion scenarios, achieving a groundbreaking synergistic enhancement between edge features and region-specific deep features. Consequently, the proposed method significantly improves the discriminative power of detection features. Extensive experiments on YOLOv8, ViT, and SSD detectors demonstrate that WaveMFFM effectively mitigates occlusion problems, thus improving the prohibited item detection performance of these representative methods.

Keywords: object detection; feature fusion; wavelet transform; prohibited item; X-ray
CLC number: TP399 **Document code**: A
Article ID: 1672-5220(2026)02-0112-08

Open Science Identity (OSID)

0 Introduction

As the population density in public transportation hubs continues to increase, the role of security checks in ensuring public safety has become increasingly important. To ensure public safety, security checks typically rely on X-ray scanning technology to detect prohibited items in passengers' luggages. However, as shown in Fig. 1, the random stacking and severe overlapping of objects in the luggage often lead to occlusion, making it more difficult to visually identify prohibited items. As a result, security personnel may struggle to accurately detect every prohibited item while processing a large number of complex X-ray images over extended periods, potentially posing risks to public safety. At the same time, frequent shift changes consume substantial human resources, making it difficult to address the issue efficiently.

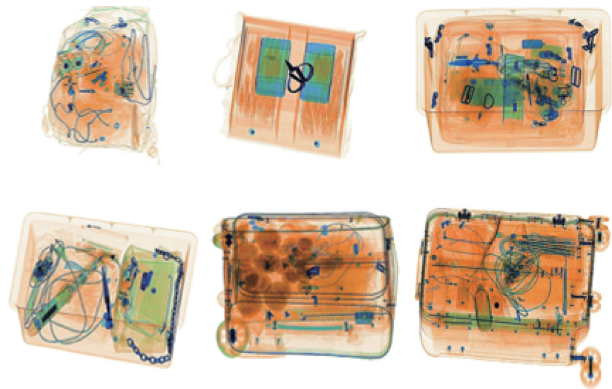


Fig. 1 Samples under X-ray security detection

Therefore, there is an urgent need for a fast, accurate, and automated method to assist security personnel in detecting prohibited items. With the rapid development of deep learning technologies, many researchers have started exploring deep learning-based solutions, especially to improve detection accuracy in cases of occlusion.

The issue of object occlusion has long been a significant challenge in prohibited item detection. Wei et al.[1] proposed an effective solution to this problem by combining color and contour information within an attention mechanism. Zhao et al.[2] introduced a label-aware mechanism that established relationships between feature channels and different labels, adjusting the features based on the assigned labels to improve detection accuracy. Shao et al.[3] introduced a foreground-background separation framework, successfully isolating prohibited items from other irrelevant objects. Ma et al.[4] employed image segmentation techniques and used dense backward connections to eliminate background interference, while optimizing prohibited item boundary extraction through an attention mechanism. Li et al.[5] adjusted the label assignment of positive samples based on the predicted intersection over union (IoU), enabling the model to focus more on extracting effective features from the foreground and reducing background interference.

Received date: 2025-01-21
* Correspondence should be addressed to CHEN Guangfeng, email: chengf@dhu.edu.cn
Citation: SUN P, CHEN G F. WaveMFFM: wavelet-guided multi-feature fusion module for X-ray prohibited item detection [J]. *Journal of Donghua University (English Edition)*, 2026, 43(2): 112-119.

Moreover, researchers have explored more efficient convolutional structures for feature extraction. Wavelet-based network designs have been shown to enhance the spatial and frequency domain representations in deep learning models, thereby improving detection performance. This technology has made significant progress in fields such as image generation[6-7], image restoration[8], and image dehazing[9], and has demonstrated strong performance in specific applications, such as underwater image detection[10] and aircraft trajectory prediction[11].

To address the occlusion problem in prohibited item detection and improve detection accuracy, this study proposes a wavelet-guided multi-feature fusion module (WaveMFFM). WaveMFFM initially processes input data through two distinct submodules to respectively extract edge features and regional depth features of prohibited items, effectively segregating targets from background interference. Subsequently, a third submodule densely integrates these extracted features with intrinsic image attributes to generate novel attention maps, thereby strengthening holistic recognition capability for prohibited items. Notably, during feature extraction, WaveMFFM employs a specifically designed de-occlusion wavelet convolution (DOWC) to resolve feature leakage caused by fixed receptive fields in conventional convolution. Through systematic separation, processing, and fusion of low-frequency and high-frequency components, DOWC constructs more robust attention maps for subsequent detection phases. WaveMFFM demonstrates enhanced performance in handling occlusions while maintaining discriminative power for subtle prohibited item features.

1 Methodology

1.1 Dataset

The performance of deep learning models is largely dependent on the quality of the dataset, as only high-quality datasets can effectively evaluate the model's detection capabilities. Therefore, a specialized dataset with high-quality annotations is crucial for both model training and evaluation. In this study, the OPIXray dataset[1] has been extended. Since the types of prohibited items should not be limited to knives, and the OPIXray dataset includes only five types of knives, namely folding knife (FO), straight knife (ST), scissor (SC), multi-tool knife (MU), utility knife (UT), its scope is too narrow, which could result in the model recognizing only knives and failing to detect other types of prohibited items. However, OPIXray's classification of occlusion levels is particularly intuitive and effective for addressing occlusion issues in prohibited item images. To overcome this limitation, the dataset is supplemented with four different prohibited items: portable charger (PO), water (WA), cosmetic (CO), and gun (GUN) from other public datasets[12-13], resulting in a new dataset for comprehensive evaluation in this study. The samples of different prohibited items are shown in Fig. 2.

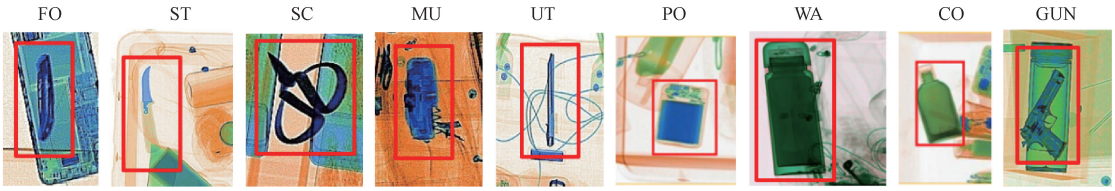


Fig. 2 Different prohibited items under X-ray

In the original OPIXray dataset, the number of samples varies significantly across different types of knives. For example, the MU has up to 2 042 images, while the ST has only 1 044 images, almost half the number of MUs. To eliminate detection bias caused by this imbalance, the number of samples for each prohibited item has been standardized to 800 images, with 600 images for training and 200 images for testing. Additionally, the occlusion level classification from the original dataset (OL1, OL2, and OL3) is retained. The category distribution of different occlusion levels in the testing set is shown in Table 1.

Table 1 Category distribution of different occlusion levels in testing sets

Testing set	Occlusion level	Number of images								
		FO	ST	SC	UT	MU	PO	WA	CO	GUN
OL1	No or slight	105	75	112	124	109	94	102	112	106
OL2	Partial	56	70	48	42	63	56	63	60	53
OL3	Severe or full	39	55	40	34	28	50	35	28	41

1.2 WaveMFFM

WaveMFFM consists of three submodules: the edge feature extraction module (EG), the region-specific deep feature extraction module (RE), and the multi-feature dense fusion module (MDFM). These submodules extract different levels of information from the input image. WaveMFFM then combines this information to generate an attention distribution map, serving as a high-quality mask for each input sample, which produces high-quality feature maps and provides

the detector with recognizable information. Unlike traditional convolution-based feature extraction, this study introduces a DOWC that excels in extracting usable features from images.

Figure 3 illustrates the overall architecture of the proposed WaveMFFM. WaveMFFM extracts different features from the image through three parallel branches and fuses them to generate a new attention distribution map, thereby enhancing the feature information for improved detection accuracy.

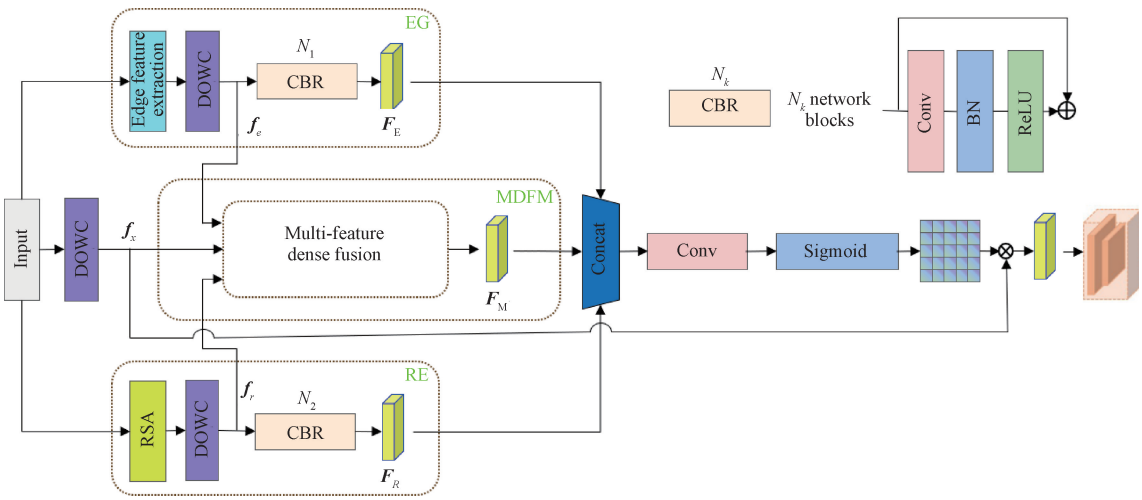
Specifically, for each input image $\boldsymbol{X}$, WaveMFFM passes it through the EG, the RE, and the MDFM, producing the corresponding feature maps: the edge feature map $\boldsymbol{F}_{\mathrm{E}}$, region-specific deep feature map $\boldsymbol{F}_{\mathrm{R}}$, and multi-feature dense fusion feature map $\boldsymbol{F}_{\mathrm{M}}$, each focusing on different aspects of the image.

In the EG branch, edge detection is first applied to extract the edge map. After processing the low-frequency and high-frequency information using DOWC, the branch obtains the temporary edge feature $\boldsymbol{f}_{\mathrm{e}}$, which is further refined through multiple CBR blocks, where CBR means convolution, batch normalization (NB), and rectified linear unit (ReLU) activation, resulting in the final edge feature map $\boldsymbol{F}_{\mathrm{E}}$. This feature map highlights the complete edges of prohibited items, especially in occluded regions.

In the RE branch, the image is first passed through the region-specific attention module (RSA) to extract the region-specific information of prohibited items. Similar to the EG branch, after DOWC processing, the branch obtains the temporary features $\boldsymbol{f}_{\mathrm{r}}$, which are then refined through multiple CBR blocks to produce the final region-specific deep feature map $\boldsymbol{F}_{\mathrm{R}}$.

In the MDFM, the intermediate features $\boldsymbol{f}_{\mathrm{e}}$ and $\boldsymbol{f}_{\mathrm{r}}$ obtained from the EG and RE branches, along with the image $\boldsymbol{X}$ processed directly through DOWC, are densely fused to generate the final feature map $\boldsymbol{F}_{\mathrm{M}}$, which integrates all key features of the prohibited items.

Finally, $\boldsymbol{F}_{\mathrm{E}}$, $\boldsymbol{F}_{\mathrm{R}}$, and $\boldsymbol{F}_{\mathrm{M}}$ are combined to generate the attention distribution map $\boldsymbol{S}$. With the help of $\boldsymbol{S}$, the module obtains enhanced features $\boldsymbol{F}$ from the input image $\boldsymbol{X}$ to facilitate accurate detection.



N_k— number of CBR blocks, $k \in \mathbf{R}$; Conv—convolution; Concat—concatenate.

Fig. 3 WaveMFFM's architecture

1.2.1 DOWC network

The network structure of DOWC is shown in Fig. 4. Specifically, DOWC uses a set of four filters to perform depth convolution with a stride of 2 to extract the low-frequency and high-frequency components of the image $\boldsymbol{X}$.

$$\begin{cases} \boldsymbol{f}_{\mathrm{LL}} = \dfrac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \\ \boldsymbol{f}_{\mathrm{LH}} = \dfrac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \\ \boldsymbol{f}_{\mathrm{HL}} = \dfrac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \\ \boldsymbol{f}_{\mathrm{HH}} = \dfrac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \end{cases} \tag{1}$$

where $\boldsymbol{f}_{\mathrm{LL}}$ is the low-pass filter; $\boldsymbol{f}_{\mathrm{LH}}$, $\boldsymbol{f}_{\mathrm{HL}}$, and $\boldsymbol{f}_{\mathrm{HH}}$ form a set of high-pass filters. For each input channel, the convolution output is

$$[\boldsymbol{X}_{\mathrm{LL}}, \boldsymbol{X}_{\mathrm{LH}}, \boldsymbol{X}_{\mathrm{HL}}, \boldsymbol{X}_{\mathrm{HH}}] = \mathrm{Conv}([\boldsymbol{f}_{\mathrm{LL}}, \boldsymbol{f}_{\mathrm{LH}}, \boldsymbol{f}_{\mathrm{HL}}, \boldsymbol{f}_{\mathrm{HH}}], \boldsymbol{X}), \tag{2}$$

where $\boldsymbol{X}_{\mathrm{LL}}$ is a low-frequency approximation that retains the main structural information of the feature map at a coarse-grained level; $\boldsymbol{X}_{\mathrm{LH}}$, $\boldsymbol{X}_{\mathrm{HL}}$, and $\boldsymbol{X}_{\mathrm{HH}}$ are the high-frequency components, which provide detailed information at a fine-grained level while preserving a large amount of noise in the feature map.

DOWC employs a cascading operation, applying two wavelet convolutions to the input. Unlike the approach reported in Refs. [14] and [15], where all high-frequency components are discarded as noise (despite containing a large amount of detailed

information), and different from the approach reported in Ref. [16], which merges processed low-frequency and high-frequency components via inverse wavelet transform, DOWC concatenates the low-frequency and high-frequency components produced by each wavelet convolution along the channel dimension, and a 1 × 1 convolution kernel is then used to compress all the information into a single feature dimension. After upsampling, the result is fused with the features before wavelet convolution. The 1-level combined operation can be given by:

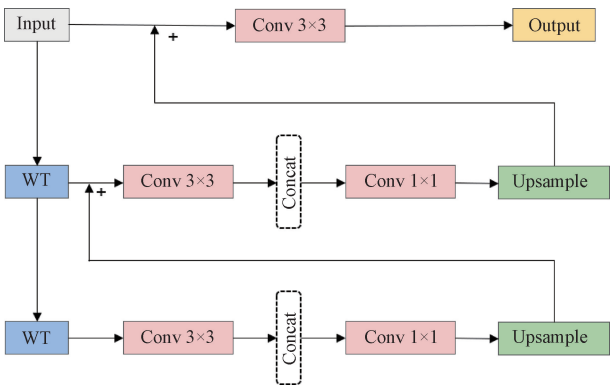
$$\boldsymbol{M}_{\mathrm{LL}} = \mathrm{W}_{c3}(X_{\mathrm{LL}}), \tag{3}$$

$$\boldsymbol{M} = \mathrm{W}_{c1}(\mathrm{Concat}(\boldsymbol{M}_{\mathrm{LL}}, \boldsymbol{X}_{ij})), \tag{4}$$

$$\boldsymbol{X}_{\mathrm{O}} = \mathrm{W}_{c3}(\boldsymbol{X}_{\mathrm{LL}} + \mathrm{upsample}(\boldsymbol{M})), \tag{5}$$

where $\mathrm{W}_{c3}(\cdot)$ is a 3 × 3 convolution; $\mathrm{W}_{c1}(\cdot)$ is a 1 × 1 convolution; $\boldsymbol{M}_{\mathrm{LL}}$ represents the processed low-frequency map; $\boldsymbol{X}_{ij}$ represents all three high-frequency maps; $\boldsymbol{M}$ represents the feature map which combines the low-frequency and high-frequency features; $\boldsymbol{X}_{\mathrm{O}}$ is the output of 1-level wavelet convolution.

In the same way, the 2-level wavelet convolution can be conducted. Through this cascading operation, DOWC achieves a significantly large receptive field without excessive parameterization, enabling faster processing while maintaining high-quality reconstruction. By extracting and processing both low-frequency and high-frequency information, DOWC generates features that are more conducive to effective detection.



WT—wavelet transform.

Fig. 4 DOWC's working principle

1.2.2 EG structure

For each input image $\boldsymbol{X}$, the edge map $\boldsymbol{E}$ of the image is generated by using a convolutional neural network with the Sobel operator. Then, by applying DOWC, the temporary edge feature $\boldsymbol{f}_{\mathrm{e}}$ is generated by using both low-frequency and high-frequency information. To make EG focus more on the edge information of prohibited items, N_1 CBR blocks are employed, where each block consists of a convolutional layer with a 3 × 3 kernel, a BN layer, and a ReLU activation, connected through residual connections to generate the enhanced edge feature map $\boldsymbol{F}_{\mathrm{E}}$. These operations can be expressed as

$$\boldsymbol{E} = \mathrm{Sobel}(\boldsymbol{X}), \tag{6}$$

$$\boldsymbol{f}_{\mathrm{e}} = \mathrm{DOWC}(\boldsymbol{E}), \tag{7}$$

$$\boldsymbol{F}_{\mathrm{E}} = \{\mathrm{ReLU}(\boldsymbol{W}_{\mathrm{e}} \cdot \boldsymbol{f}_{\mathrm{e}} + \boldsymbol{b}_{\mathrm{e}})\}_{N_1}, \tag{8}$$

where $\{\cdot\}_{N_1}$ indicates that the operation is repeated N_1 times; $\boldsymbol{W}_{\mathrm{e}}$ and $\boldsymbol{b}_{\mathrm{e}}$ are the parameters of the convolutional layers.

1.2.3 RE structure

The region-specific depth features essentially represent the attention distribution mask of the input image, where each value indicates the importance of the corresponding pixel in the image. This method effectively eliminates the impact of color distribution differences, significantly enhances the information of the prohibited item regions, and preserves the geometric structure of the image. It successfully differentiates prohibited items from the background and other objects, thereby improving the model's generalization capability.

RSA adopts the attention mechanism of the convolutional block attention module[17] (as shown in Fig. 5). After normalization and standardization preprocessing, the input image passes through two convolutional layers to generate a feature map $\boldsymbol{F}_1$ (with dimensions $W/2 \times H/2$ and 32 channels). $\boldsymbol{F}_1$ is then input to the channel attention module (CAM), which outputs the channel attention mask $\boldsymbol{A}_{\mathrm{c}}$ (with dimensions of 1×1 and 32 channels), representing the weight coefficients of each channel feature. Next, $\boldsymbol{A}_{\mathrm{c}}$ is element-wise multiplied with $\boldsymbol{F}_1$ along the corresponding channels to obtain the new feature map $\boldsymbol{F}_2$. $\boldsymbol{F}_2$ is then passed into the spatial attention module (SAM), which outputs the spatial attention mask $\boldsymbol{A}_{\mathrm{s}}$ (with dimensions of $0.5W \times 0.5H$ and 1 channel). This mask is upsampled by using bilinear interpolation to match the size of the input image, resulting in the intermediate feature map $\boldsymbol{F}_{\mathrm{I}}$. Similar to the processing in EG, $\boldsymbol{F}_{\mathrm{I}}$ is further processed by DOWC, followed by adaptive learning with N_2 CBR network blocks, ultimately obtaining the regional depth feature map $\boldsymbol{F}_{\mathrm{R}}$.

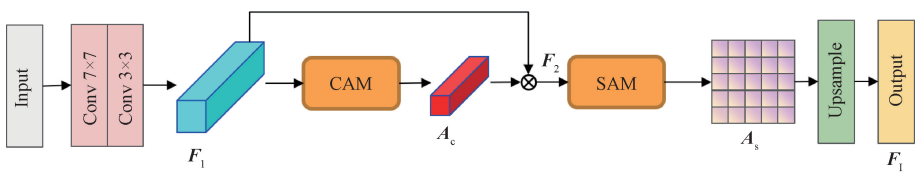
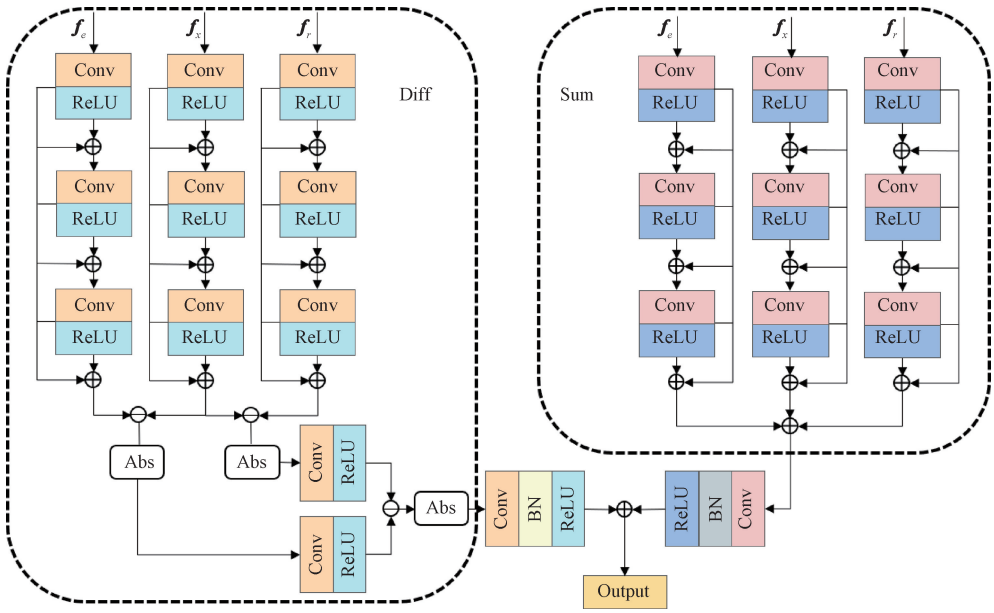


Fig. 5 RSA's working principle

1.2.4 MDFM structure

As shown in Fig. 6, the MDFM structure consists of two branches: the sum branch and the difference branch. The sum branch fuses the edge feature f_e and the regional depth feature f_r extracted from EG and RE, along with the overall feature f_o obtained by applying DOWC to the image, through an addition operation. This fusion helps enhance the overall information of the prohibited items in the image. The difference branch calculates the difference between f_o and f_e, as well as f_r, thus facilitating the separation of prohibited items from the background. Each branch consists of three densely connected streams with shared weights, where all convolution operations use a 3×3 kernel.



Diff—the difference branch; Sum—the sum branch; Abs—absolute.

Fig. 6 MDFM's procedure

The purpose of the MDFM design is to integrate multiple features within each stream, providing the model with more comprehensive information for improved decision-making. This structure not only increases the robustness of the model but also enhances feature correction due to the rich residual connections in the dense connections, thereby generating more discriminative feature maps that significantly improve the performance of prohibited item detection.

2 Experiment and Analysis

In this section, extensive experiments were conducted to comprehensively evaluate the proposed WaveMFFM. Firstly, this section validated the superiority of WaveMFFM over other detection methods across different categories and occlusion levels. Secondly, ablation studies were performed to systematically assess the contribution and effectiveness of each module within WaveMFFM. Finally, this study examined the generalizability of WaveMFFM across multiple detectors and verified the performance improvement when WaveMFFM was integrated into existing detection models.

This study was conducted on a cloud computing platform with the following configurations. The hardware environment comprised an Intel Xeon (R) Gold 6430 CPU and an NVIDIA RTX 4090 GPU (24 GB). The software setup included PyTorch 2.1.0 and Python 3.10 under Ubuntu 22.04. All experiments utilized the extended OPIXray dataset constructed in Subsection 1.1. Model performance was evaluated by using the mean average precision (mAP) metric for object detection, with the IoU threshold set to 0.5.

2.1 Comparison of different detection methods

In this part, YOLOv8[18] was used as the base detection model and compared the performance of the proposed WaveMFFM with three other detection methods, SE[19], DOMA[1], and FEM[20]. The experimental results are recorded in Tables 2 and 3.

As shown in Table 2, the YOLOv8 detector integrated with WaveMFFM outperforms the baseline YOLOv8 model and those integrated with SE, DOMA, and FEM in terms of detection performance. Specifically, the mAP improved by 1.85, 1.22, 0.82, and 0.92 percentage points, respectively. Notably, compared with the baseline YOLOv8 model, the detection performance for low-precision categories (ST and CO) saw the most significant improvements, with detection accuracy increasing by 2.72 and 3.73 percentage points, respectively. These results demonstrate that WaveMFFM effectively enhances YOLOv8's performance across various detection tasks.

Table 2 Comparison of different detection methods for each item

Method	mAP/%									
	Average	FO	ST	SC	UT	MU	PO	WA	CO	GUN
YOLOv8	78.73	83.52	60.52	95.58	72.31	85.32	87.51	84.9	42.13	96.74
YOLOv8+SE	79.36	83.6	61.51	95.66	72.29	85.97	88.35	84.93	44.59	97.36
YOLOv8+DOMA	79.76	83.93	61.72	95.97	72.55	86.9	88.72	85.32	45.43	97.27
YOLOv8+FEM	79.66	84.12	62.35	95.45	72.63	86.58	88.39	85.24	44.96	97.22
YOLOv8+WaveMFFM	**80.58**	**84.95**	**63.24**	**96.21**	**73.47**	**87.35**	**89.11**	**86.58**	**45.86**	**98.43**

Table 3 Comparison of different detection methods at different occlusion levels

Method	mAP/%		
	OL1	OL2	OL3
YOLOv8	82.43	79.26	74.50
YOLOv8+SE	83.05	79.57	75.46
YOLOv8+DOMA	83.33	80.18	75.77
YOLOv8+FEM	83.52	79.86	75.60
YOLOv8+WaveMFFM	**84.35**	**80.31**	**77.08**

As observed in Table 3, WaveMFFM consistently achieves the highest detection accuracy across all occlusion levels. When compared with the baseline YOLOv8, the integration of WaveMFFM leads to an increase in detection accuracy by 1.92, 1.05, and 2.58 percentage points at occlusion levels OL1, OL2, and OL3, respectively. These findings further confirm that WaveMFFM exhibits strong adaptability and effectively handles the challenges posed by object occlusion in the detection of different types of prohibited items.

2.2 Ablation study

To comprehensively evaluate the contribution of each component of WaveMFFM, YOLOv8 was used as the baseline detector and compared with a method where the input image is directly fed into the detector without any additional processing. As shown in Table 4, the experimental results show that when EG, RE, and MDFM are sequentially incorporated, the detection accuracy increases by 0.58, 0.70, and 1.02 percentage points, respectively. When all three components are combined, the detection accuracy improves by 1.85 percentage points. These results confirm that each module of WaveMFFM effectively extracts features and significantly enhances the overall performance of the detector.

Table 4 Performance of each part of WaveMFFM

Method	mAP/%									
	Average	FO	ST	SC	UT	MU	PO	WA	CO	GUN
YOLOv8	78.73	83.52	60.52	95.58	72.31	85.32	87.51	84.9	42.13	96.74
YOLOv8+EG	79.31	83.63	61.52	95.33	72.22	86.68	88.16	85.63	43.46	97.13
YOLOv8+RE	79.43	83.94	61.58	95.39	72.4	86.28	88.23	85.21	44.56	97.31
YOLOv8+MDFM	79.75	83.56	62.29	95.53	72.39	86.85	89.01	85.52	44.86	97.75
YOLOv8+WaveMFFM	**80.58**	**84.95**	**63.24**	**96.21**	**73.47**	**87.35**	**89.11**	**86.58**	**45.86**	**98.43**

To validate the feature extraction capability of DOWC over conventional convolution structures, DOWC was replaced in WaveMFFM with a standard convolution structure while ensuring that the output feature map size remained unchanged. The experimental results, as shown in Table 5, indicate that the detection accuracy using DOWC for feature extraction significantly outperforms that of the conventional convolution structure across various occlusion levels. Specifically, at occlusion levels OL1, OL2, and OL3, the detection accuracy improves by 0.76, 0.26, and 0.96 percentage points, respectively. The most significant improvement in detection accuracy is observed at the highest occlusion level (OL3).

Table 5 Performance of DOWC

Method	mAP/%		
	OL1	OL2	OL3
YOLOv8+WaveMFFM (normal)	83.59	80.05	76.12
YOLOv8+WaveMFFM (with DOWC)	84.35	80.31	77.08

2.3 Comparison of different detectors

To further evaluate the effectiveness of WaveMFFM and verify its scalability, WaveMFFM was integrated into the SSD[21], YOLOv8, and ViT[22] detection models and conducted comparative experiments. The experimental results are shown in Table 6. After integrating WaveMFFM, the detection accuracy increased by 3.90,

1.85, and 2.45 percentage points, compared with the original models (SSD, YOLOv8, and ViT), respectively. These results demonstrate that WaveMFFM can be used as a plug-and-play component, successfully embedded into most detectors, and significantly improve their performance.

Table 6 Performance of WaveMFFM in different detectors

Method	mAP/%									
	Average	FO	ST	SC	UT	MU	PO	WA	CO	GUN
SSD	71.37	76.92	35.1	93.45	65.78	83.12	80.19	81.25	36.21	90.29
SSD+WaveMFFM	**75.27**	**83.52**	**42.74**	**96.24**	**69.61**	**84.53**	**83.14**	**84.5**	**39.43**	**93.68**
YOLOv8	78.73	83.52	60.52	95.58	72.31	85.32	87.51	84.9	42.13	96.74
YOLOv8+WaveMFFM	**80.58**	**84.95**	**63.24**	**96.21**	**73.47**	**87.35**	**89.11**	**86.58**	**45.86**	**98.43**
ViT	76.57	80.13	49.25	96.76	69.48	85.37	85.58	84.72	41.32	96.55
ViT+WaveMFFM	**79.02**	**82.89**	**54.77**	**97.12**	**73.23**	**86.57**	**89.65**	**86.23**	**43.76**	**96.98**

3 Conclusions

This study proposes a wavelet-guided multi-feature fusion module, named WaveMFFM. For the first time, the wavelet transform is introduced into the prohibited item detection task, and DOWC has been designed. As a plug-and-play module, WaveMFFM can be easily embedded into various detectors to enhance their performance. Through comprehensive evaluation, experimental results demonstrate that WaveMFFM significantly addresses occlusion issues in prohibited item detection and effectively improves detection accuracy. Comparative experiments with other existing methods further validate the superiority of WaveMFFM, highlighting its broad potential for practical applications.

References

[1] WEI Y L, TAO R S, WU Z J, et al. Occluded prohibited items detection: an X-ray security inspection benchmark and de-occlusion attention module [C]//Proceedings of the 28th ACM International Conference on Multimedia. New York: Association for Computing Machinery, 2020: 138-146.

[2] ZHAO C R, ZHU L, DOU S G, et al. Detecting overlapped objects in X-ray security imagery by a label-aware mechanism [J]. *IEEE Transactions on Information Forensics and Security*, 2022, 17: 998-1009.

[3] SHAO F T, LIU J, WU P, et al. Exploiting foreground and background separation for prohibited item detection in overlapping X-ray images [J]. *Pattern Recognition*, 2022, 122: 108261.

[4] MA B W, JIA T, SU M, et al. Automated segmentation of prohibited items in X-ray baggage images using dense de-overlap attention snake [J]. *IEEE Transactions on Multimedia*, 2023, 25: 4374-4386.

[5] LI M Y, MA B W, WANG H, et al. GADet: a geometry-aware X-ray prohibited items detector [J]. *IEEE Sensors Journal*, 2024, 24(2): 1665-1678.

[6] YANG M P, WANG Z, CHI Z Q, et al. WaveGAN: frequency-aware GAN for high-fidelity few-shot image generation [C]// European Conference on Computer Vision. Cham: Springer, 2022: 1-17.

[7] ZHANG B W, GU S Y, ZHANG B, et al. StyleSwin: transformer-based GAN for high-resolution image generation [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2022: 11294-11304.

[8] HSU W Y, JIAN P W. Detail-enhanced wavelet residual network for single image super-resolution [J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 1-13.

[9] HWANG S, HAN D, JUNG C, et al. WaveDH: wavelet sub-bands guided ConvNet for efficient image dehazing[EB/OL]. (2024-04-02)[2024-12-01]. https://arxiv.org/abs/2404.01604.

[10] ZHAO C, CAI W L, DONG C Y, et al. Wavelet-based Fourier information interaction with frequency diffusion adjustment for underwater image restoration [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2024: 8281-8291.

[11] ZHANG Z, GUO D Y, ZHOU S Z, et al. Flight trajectory prediction enabled by time-frequency wavelet transform [J]. *Nature Communications*, 2023, 14: 5258.

[12] MIAO C J, XIE L X, WAN F, et al. SIXray: a large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition

(CVPR). New York: IEEE, 2019: 2114-2123.
[13] TAO R S, WEI Y L, JIANG X J, et al. Towards real-world X-ray security inspection: a high-quality benchmark and lateral inhibition module for prohibited items detection[EB/OL]. (2021-08-01) [2025-01-01]. https://arxiv.org/abs/2108.09917.
[14] LI Q F, SHEN L L, GUO S, et al. WaveCNet: wavelet integrated CNNs to suppress aliasing effect for noise-robust image classification[J]. *IEEE Transactions on Image Processing*, 2021, 30: 7074-7089.
[15] WILLIAMS T, LI R Y. Wavelet pooling for convolutional neural networks[C]//International Conference on Learning Representations. [S.l.]: OpenReview.net, 2018.
[16] FINDER S E, AMOYAL R, TREISTER E, et al. Wavelet convolutions for large receptive fields [C]//Computer Vision-ECCV 2024. Cham: Springer, 2025: 363-380.
[17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]// Computer Vision-ECCV 2018. Cham: Springer, 2018: 3-19.
[18] VARGHESE R, SAMBATH M. YOLOv8: a novel object detection algorithm with enhanced performance and robustness [C]//2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). New York: IEEE, 2024: 1-6.
[19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 7132-7141.
[20] ZHANG Y, YE M, ZHU G Y, et al. FFCA-YOLO for small object detection in remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-15.
[21] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C]// Computer Vision-ECCV 2016. Cham: Springer, 2016: 21-37.
[22] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. (2020-10-22) [2024-12-01]. https://arxiv.org/abs/2010.11929.

WaveMFFM：用于 X 射线违禁品检测的小波变换引导的多特征融合模块

孙 鹏，陈广锋*
东华大学 机械工程学院，上海 201620

摘 要：为了更准确地检测 X 射线图像中的违禁品，提出了一种小波变换引导的多特征融合模块(wavelet-guided multi-feature fusion module，WaveMFFM)，其作为即插即用组件，可轻松集成至现有检测器中。WaveMFFM 引入小波变换，构建了去遮挡小波卷积(de-occlusion wavelet convolution，DOWC)结构，通过频域解耦机制将低频全局轮廓信息与高频细节纹理特征进行动态融合，有效解决了现有卷积操作在遮挡场景下的特征混淆问题，实现了外形边缘特征与区域深度特征的协同增强，显著提升了检测特征的有效性。在 YOLOv8、ViT 和 SSD 检测器上的试验表明，WaveMFFM 有效缓解了遮挡问题，从而提升了这些代表性方法的违禁品检测性能。

关键词：目标检测；特征融合；小波变换；违禁品；X 射线