A Terahertz Image Target Recognition Method Based on Improved YOLOv5

Juan Chen , Zhiqiang Yang * , Wanjun Wang , Lei Gong , Lihong Yang , Liguo Wang , Yao Li , Yan Feng

Posted Date: 22 September 2025

doi: 10.20944/preprints202509.1760.v1

Keywords: terahertz image; target recognition; deep learning; YOLOv5; attention mechanism

 doi:10.20944/preprints202509.1760.v1

Article

A Terahertz Image Target Recognition Method Based on Improved YOLOv5

Juan Chen [1], Zhiqiang Yang [1,*], Wanjun Wang [1], Lei Gong [1], Lihong Yang [1], Liguo Wang [1], Yao Li [1] and Yan Feng [2]

[1] School of Photoelectric Engineering, Xi'an Technological University, Xi'an 710021, China

[2] North Electronic-Optics Co., Lit, Xian 710043, China

* Correspondence: yangzhiqiang@xatu.edu.cn

Abstract

Terahertz (THz) waves exhibit a number of advantageous properties, including low radiation, and penetration capability, in addition to a broad bandwidth. These characteristics render terahertz image target recognition a prominent area of research. To address the issue of insufficient recognition accuracy of terahertz images when targets are blurred and features are limited, we introduce Enhanced and Occlusion-aware Focus YOLOv5(EOF-YOLOv5), an improved model based on YOLOv5. This study performs image enhancement preprocessing on raw terahertz image datasets acquired by a terahertz near-field array imaging system to enhance target contrast and contour information. For the improved network EOF-YOLOv5, first, an Occlusion-Aware Context Attention (OCA) mechanism is integrated into the neck network of YOLOv5, which employs deformable convolution to predict occluded regions, dynamically adjusts attention weights, and enhances feature responses in visible areas. Second, the original Complete Intersection over Union (CIoU) loss function was replaced with the Focal-Efficient Intersection over Union (Focal-EIoU) loss function to mitigate excessive focus on simple samples and enhance the learning capability for challenging samples. The experimental results demonstrate that image enhancement preprocessing enhances both visual quality and structural information while improving network recognition accuracy. On the same preprocessed dataset, the EOF-YOLOv5 algorithm achieved a 12.6% increase in precision (P) to 79.3%, a 7.5% improvement in recall (R) to 80.6%, an 8.2% boost in mean average precision (mAP50) to 83.7% compared to the baseline YOLOv5 model. Compared with other mainstream algorithms, the proposed algorithm achieves more accurate detection of object locations, validating its feasibility and effectiveness.

Keywords: terahertz image; target recognition; deep learning; YOLOv5; attention mechanism

1. Introduction

Terahertz waves are defined as a segment of the electromagnetic spectrum, characterized by frequencies ranging from 0.1 THz to 10 THz, with corresponding wavelengths spanning 0.03 to 3 millimeters [1]. Terahertz waves are located between infrared and microwave spectra, combining the penetrating power of microwaves with the resolution and biological safety of infrared. In low-humidity atmospheric environments or short-distance conditions, they can achieve high-resolution imaging. With the development of technologies such as focal plane arrays, real-time high-frame-rate terahertz imaging has also become possible [2–4]. This technology demonstrates extensive application potential in both military and civilian domains.

In the context of contemporary information systems, target recognition technology plays a pivotal role in reducing uncertainties in situational awareness, thereby facilitating rapid and accurate target identification. Advancements in terahertz imaging algorithms, driven by continuous iterations and refinements, have led to substantial advancements in imaging clarity. These advancements have established a robust foundation for target recognition, paving the way for significant progress in the

 doi:10.20944/preprints202509.1760.v1

field. The conventional approach to terahertz image target recognition entails the utilisation of manual feature extraction techniques, including threshold segmentation and edge detection [5,6]. However, this approach is constrained by inherent limitations in terahertz imaging, including texture deficiency, blurred contours, and background interference, resulting in cumbersome processing procedures and suboptimal detection accuracy. In recent years, with the advent of deep learning, neural network-based object detection algorithms have emerged as the prevailing approach, which can be broadly categorized into two methodologies: single-stage (One-stage) and two-stage (Two-stage) frameworks. The two-stage object detection algorithm initially employs region proposal generation techniques to produce candidate regions potentially containing targets, followed by position regression and category prediction for these candidate regions. The representative methodologies comprise Fast R-CNN, Faster R-CNN, and analogous algorithms [7]. The elimination of region proposal extraction in single-stage object detection results in the acquisition of both object categories and locations through a single forward pass. Prominent single-stage methodologies currently include SSD [8], RetinaNet [9], and the YOLO (You Only Look Once) series [10–15]. In the current stage, YOLO has emerged as the preferred choice for object detection algorithms due to its rapid inference speed and compact model size. H. Xiao et al. proposed a rapid terahertz image detection algorithm that integrates preprocessing techniques with structural optimization of regional convolutional neural networks, achieving highly efficient detection performance in dense flow scenarios [16]. C. Li et al. proposed a terahertz image detection algorithm based on spatiotemporal information, which comprises two main components: coarse detection and fine recognition. This algorithm fully leverages the spatiotemporal information contained in terahertz security image sequences, along with the efficient feature extraction capabilities of neural networks, leading to a significant improvement in detection efficiency [17]. M. Kovbasa et al. proposed a detection algorithm for postal terahertz scanners, that identifies concealed objects at 0.14 THz by incorporating convolutional neural networks into the scanning system [18]. Xu et al. developed a multi-scale filtering geometric enhancement method that incorporates spatial distance grids derived from geometric transformation matrices, significantly improving the detection accuracy of convolutional neural networks (CNN) for passive terahertz images [19]. Danso et al. introduced the concept of transfer learning based on the YOLOv5 model, improving the accuracy of defect detection in terahertz images [20]. Yu et al. proposed the BoT-YOLO+ model based on the YOLOv5 network architecture. This model enhances the feature extraction capabilities for component overlaps and small targets in terahertz radar images by introducing a BoTNet backbone network that integrates Transformer attention mechanisms [21]. Ge et al. proposed a lightweight algorithm based on YOLOv7 (LWMD-YOLOv7), which integrates the SPD-Mobile network and the large selective kernel (LSK) module to address the issues of low image quality and real-time detection in terahertz security screening [22]. Zeng et al. proposed a covert hazardous object detection method for terahertz images based on YOLOv8, called Cross-Feature Fusion Transformer YOLO (CFT-YOLO). It achieves cross-channel and cross-spatial information transfer between feature maps, enhancing the perception of target objects [23]. Yang et al. proposed a method for detecting parabolic antennas in satellite inverse synthetic aperture radar images in the terahertz domain based on component prior knowledge and an improved YOLOv8 network, addressing challenges related to component identification in satellite ISAR images [24].Cheng et al. proposed an improved YOLOv8 network combining an adaptive context-aware attention network (ACAN) and a dynamic adaptive convolution block (DACB), which effectively improved the detection accuracy of hidden objects in terahertz security images [25].

The imaging of targets is predominantly characterized by contour features, with a notable absence of detailed texture information. This limitation stems primarily from the intensity constraints of terahertz wave sources and the inherent reflectivity properties of target materials. Consequently, recognition accuracy is often compromised, manifesting as a high incidence of false positives and missed detections [22]. Consequently, the effective utilization of limited features within low-resolution terahertz images to achieve rapid and precise target detection remains a critical and unresolved technical challenge in the field of terahertz image recognition. Given the limited sample

 doi:10.20944/preprints202509.1760.v1

diversity in existing terahertz image datasets, this study introduces synthetically occluded target samples to increase data variability and complexity, thereby validating the superior recognition performance of the proposed model in terahertz image analysis tasks. The occlusion of target objects substantially diminishes the available feature information in the visible regions, resulting in a degradation phenomenon that severely compromises the feature extraction capability of the network model. This makes it challenging to learn robust and discriminative features from the limited visible parts, subsequently affecting the performance of detection, recognition, or tracking tasks.

Aiming at the above problems, we carry out the following work:

(1) A terahertz near-field imaging system with planar array configuration was established, utilizing military target models including tanks, armored vehicles, and jeeps as research subjects to acquire a comprehensive terahertz image dataset.

(2) Due to the indistinct edge contours and low contrast of targets in terahertz images, an enhanced preprocessing approach is applied. This method employs an adaptive stationary wavelet transform to decompose the image into low-frequency and high-frequency components. Adaptive coefficient enhancement and multi-directional filtering fusion are then applied to the respective components. Finally, an inverse transform is performed to reconstruct the image, thereby enhancing the target contrast and contour delineation.

(3) Based on the YOLOv5 benchmark model, an occlusion-aware attention mechanism is integrated prior to the Cross Stage Partial Network with 3 convolutions (C3) module in the original YOLOv5 neck network. This mechanism consists of a channel attention branch, which extracts attention cues across channel dimensions, and an occlusion-aware branch that processes spatial dimension occlusion information. By integrating these components, the mechanism effectively suppresses background interference while enhancing feature responses in visible target regions.

(4) The original loss function Complete Intersection over Union (CIoU) is replaced with Focal-Efficient Intersection over Union (Focal-EIoU) to enhance the measurement accuracy of similarity between predicted bounding boxes and ground truth bounding boxes.

Experimental results demonstrate that the proposed EOF-YOLOv5 algorithm not only exhibits superior object recognition capabilities compared to the original YOLOv5 framework but also significantly outperforms other mainstream detection algorithms in terms of accuracy on terahertz image recognition tasks.

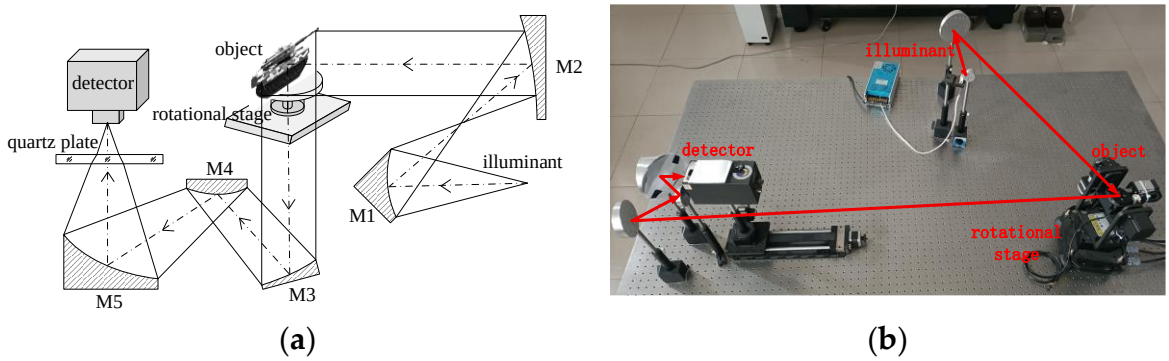
2. Materials and Methods

2.1. Terahertz Image Collection and Dataset Construction

The terahertz image dataset utilized in this study was acquired through a planar array imaging system employing terahertz near-field imaging technology. The system is constructed on a 1.5m*2m optical platform, with its structural configuration illustrated in Figure 1 (a). The terahertz detector selected is the RIGI-M2 Uncooled FPA THz micro-bolometer array camera from Tera Sense Company, which weighs less than 200g, is easy to carry and use, and does not require a complex cryogenic cooling system. It maintains high sensitivity and high frame rates while offering a low cost. Given the advantages of thermal radiation terahertz sources, including simple structure, portability, and a wide terahertz frequency spectrum range, a 100W tungsten lamp was selected as the terahertz source. It primarily composed of two key components: the terahertz collimated illumination system and a terahertz imaging reception system. Due to the low radiation efficiency of tungsten lamps in the terahertz band, quartz glass plates are selected as low-pass filters to block wavelengths between 15μm and 30μm, ensuring signal purity. In the terahertz irradiation system, the waves emitted by the tungsten lamp are collimated via a total reflection-based optical path and expanded into a beam with a diameter of 100 mm and a divergence angle of ≤0.284°, thereby effectively minimizing optical energy loss. The terahertz imaging reception system utilizes an off-axis three-mirror anastigmat configuration with a focal length of 122.4 mm and a field of view of 2.545° × 1.909°. This design achieves diffraction-limited planar array imaging through wavefront error control within $\lambda/4$. The

 doi:10.20944/preprints202509.1760.v1

terahertz imaging system achieves clear planar array imaging over an area of 100*100mm², with an imaging resolution of 0.19mm. The final configuration is illustrated in Figure 1 (b). The imaging parameters for the images are presented in Table 1.



(**a**) (**b**)

Figure 1. (**a**) Optical path schematic of the Terahertz imaging system; (**b**) Physical implementation of Terahertz optical path.

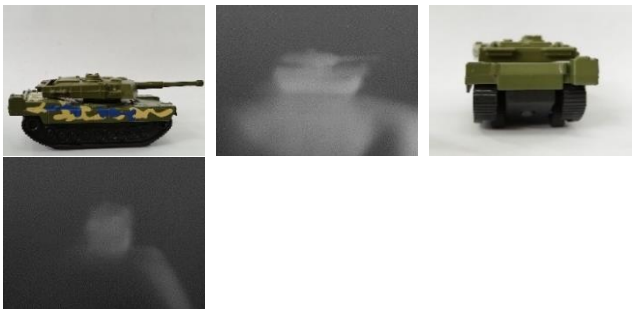
Table 1. System design parameters.

Project	parameters
Spectral Range	30~60μm
Terahertz Frequency	5~10THz
Working Distance	1500mm
Focal Length	122.44mm
Field of View	1.909°×2.545°
F Number	1.22
Number of Pixels	640×480
Detector Operating Frequency	0.1~20THz
Detector NEP	<1.5 pW/√Hz at 4.6THz
Detector Integration Time	50 μs
Quartz Plate Dimensions	30mm×30mm×1mm

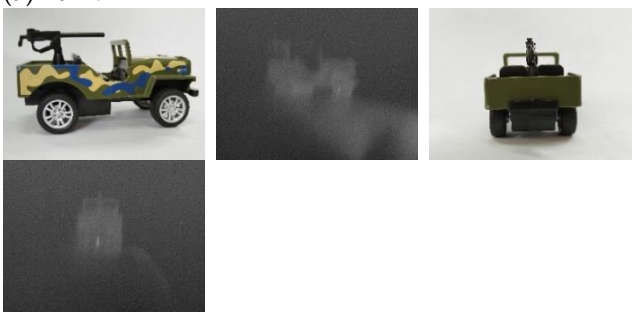
To construct a comprehensive dataset encompassing multiple perspectives and orientations of target models, a rotating platform was employed to sequentially position three distinct models—a tank, an armored vehicle, and a jeep—for image acquisition. The main body of the model car shell is made of metal (alloy material), while the gun barrel and machine gun components are plastic. The platform was rotated in 6° increments, with image captured at each interval. To enhance dataset diversity and simulate the multi-pose characteristics of vehicles under real-world conditions, the pitch angle of each model was systematically adjusted. A total of 1,618 images were captured at 6° intervals, constituting a comprehensive dataset that covers a full 360° horizontal field of view and pitch angles ranging from –30° to +30°. The dataset includes three target categories: tanks, armored vehicles, and jeeps. Representative sample images from both the visible light and corresponding terahertz datasets are presented in Figure 2.

To facilitate effective model training and validation, the 1618 images were partitioned into training and test sets according to the presence or absence of inter-object occlusion. The training set primarily consists of images containing only a single target category. The test set is composed of those images featuring two target categories with mutual occlusion. This distribution leads to a natural imbalance in the test set size across categories, as shown in Table 2, because occlusion scenarios do not occur uniformly for all object types. We introduced samples simulating target occlusion to construct a more challenging test set, enhancing the complexity of evaluation data and comprehensively testing the model's robustness. The dataset specifications are detailed in Table 2.

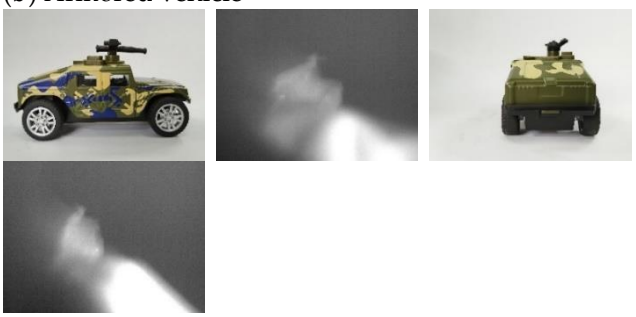
 doi:10.20944/preprints202509.1760.v1



(**a**) Tank



(**b**) Armored vehicle



(**c**) Jeep

Figure 2. presents a comparative visualization of the visible light image and its corresponding terahertz image dataset samples.

Table 2. The composition of Terahertz Dataset.

Category	Training Quantity (Units)	Test Quantity (Units)	Label Name
Tank	697	45	TK
Armored vehicle	529	81	ZJC
Jeep	495	126	JPC
Total	1721	252	/

2.2. Pre-Processing Methods for Terahertz Image

Affected by the limited intensity of terahertz wave sources and the inherent reflectivity of target materials, terahertz images are characterized by low contrast, blurred target edges, and attenuated detail features. Furthermore, due to the inconsistent brightness levels of targets in terahertz images, backgrounds with similar luminance can obscure target characteristics when their intensity values closely match those of the target. To address this issue, target features must be accentuated and contrast enhanced within the image to facilitate more efficient detection and recognition by the model. This study proposes a preprocessing method for enhancing terahertz near-field images. The image is decomposed into low-frequency and high-frequency components through stationary wavelet transform (SWT) [26]. The low-frequency component is enhanced using adaptive coefficient scaling, while the high-frequency component is refined through multi-directional filtering. Finally,

 doi:10.20944/preprints202509.1760.v1

an inverse transform is applied to reconstruct the image, thereby improving both target contrast and contour delineation.

Initially, the image $I(x, y)$ is decomposed using the Haar wavelet basis function into a low-frequency component (*LL*), which encapsulates illumination information, and high-frequency components (*LH*/*HL*/*HH*), which represent detailed features [27,28]. For the high-frequency components, a filter bank consisting of eight directionally sensitive operators is constructed to comprehensively capture anisotropic features while mitigating the directional information loss associated with single-filter approaches [29]. This ensemble incorporates horizontal, vertical, diagonal, Laplacian, and sharpening filter operators, with anisotropic feature extraction accomplished via a mean fusion strategy. For the low-frequency components, a triple statistical model— based on image mean, local standard deviation, and global standard deviation—is established. This model incorporates α-linear gain to synergistically optimize both dark region enhancement and contrast adaptation. The formulas for calculating the mean μ_{LL} and standard σ_{LL} deviation of an image are as follows:

$$\mu_{LL} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} LL(i, j) \tag{1}$$

$$\sigma_{LL} = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} (LL(i, j) - \mu_{LL})^2} \tag{2}$$

Here, M and N represent the number of rows and columns of the image, respectively, and (i, j) denote the coordinates of the image pixels.

When the image exhibits underexposure or overexposure, an adaptive adjustment mechanism based on mean intensity is employed. This mechanism utilizes a threshold-regulated approach with an amplitude coefficient of a to modulate the enhancement process. The threshold is set at 0.5, which represents the perceptual middle-gray point in the normalized L channel. When the value of μ_{LL} falls below this predetermined threshold, a positive intervention is implemented to elevate the mean value; conversely, when μ_{LL} exceeds the threshold, a negative intervention is applied to reduce the mean value. Similarly, the standard deviation of the image is adjusted based on an empirically set predefined threshold of 0.2 to detect low-contrast images, with an amplitude coefficient of b. The standard deviation is utilized to adjust the global contrast of the image, with a lower threshold of 0.15 set for detecting extremely low-contrast scenarios. The amplitude coefficient c facilitates moderate adjustments, as specified in the following formulae:

$$\alpha_{mean} = 1.0 + a \times (0.5 - \mu_{LL}) \tag{3}$$

$$\alpha_{std} = 1.0 + b \times (0.2 - \sigma_{LL}) \tag{4}$$

$$\alpha_{global} = 1.0 + c \times (0.15 - \sigma_{LL}) \tag{5}$$

In this configuration, parameter a is set to 0.8, parameter a is set to 0.8, parameter b is assigned a value of 1.5, and parameter c is initialized at 1.2.

The derived adaptive adjustment factors (mean α_{mean}, standard deviation α_{std}, and global statistic α_{global}) are integrated with the default values to achieve a balance between automated optimization and manual control. The dynamic computation enhancement factor α_{final} is determined by setting coefficient $\alpha_{adaptive}$ to 0.7, indicating that adaptive dominance accounts for 70%, prioritizing data-driven dynamic adjustments. Coefficient $\alpha_{default}$ is set to 0.3, signifying that manual intervention constitutes 30%, thereby preventing excessive correction in extreme scenarios and enhancing robustness.

 doi:10.20944/preprints202509.1760.v1

$$\alpha_{adaptive} = 0.4\alpha_{mean} + 0.3\alpha_{std} + 0.3\alpha_{global} \tag{6}$$

$$\alpha_{final} = 0.7\alpha_{adaptive} + 0.3\alpha_{default} \tag{7}$$

The derived dynamic computational enhancement coefficient α_{final} is utilized to adjust the contrast of low-frequency components:

$$LL_{enhanced} = \alpha \bullet LL \tag{8}$$

The adaptive range of α_{final} is constrained within the interval [0.7, 1.5].

The processed low-frequency and high-frequency components are synthesized via inverse wavelet transform to reconstruct the enhanced image Irec. To further optimize local contrast, Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied [30]. This technique enhances details in both dark and bright regions while mitigating the over-enhancement issues commonly associated with traditional histogram equalization.

2.3. Improved YOLOv5 Algorithm

The YOLO (You Only Look Once) series has undergone rapid development in recent years. However, given the relatively limited scale of the dataset used in this study, the sophisticated mechanisms introduced in newer versions such as YOLOv10 and YOLOv12— including dynamic label assignment—may require larger volumes of data for effective training, increasing the risk of overfitting on smaller datasets [31,32]. YOLOv5 features a relatively compact parameter size, making it more suitable for training with limited datasets. Furthermore, since terahertz images are exclusively grayscale and lack chromatic information, the multi-level feature extraction mechanisms of more complex models may introduce redundant. The streamlined architecture of YOLOv5 proves sufficiently effective in capturing essential contour variations under these conditions. Consequently, YOLOv5 was selected as the baseline model for the experimental investigations conducted in this study.

As illustrated in Figure 3, the YOLOv5 architecture consists of three main components: the Backbone, the Neck, and the Head. The Backbone is responsible for feature extraction and progressively reduces the spatial dimensions of the feature maps. The backbone primary consist of a standard convolution (Conv) module, a Cross Stage Partial Network with 3 convolutions (C3) module, and a Spatial Pyramid Pooling Fusion (SPPF) module. The Conv module is employed for downsampling feature map, adjusting channel dimensions, and performing normalization and nonlinear activation. The C3 module, composed of three Conv modules and one Bottleneck module, serves as a critical feature extraction component. The SPPF module, which stands for Spatial Pyramid Pooling, is employed to map feature maps of varying scales to a uniform scale, thereby facilitating the processing of targets of different sizes. The Neck adopts a Feature Pyramid Network (FPN) structure that effectively integrates shallow spatial features from the Backbone with deep semantic features. The Head consists of a Detect module containing three parallel 1×1 convolutional layers, each corresponding to one of the three detection feature layers. Based on the spatial dimensions of the feature maps, the model predicts objects at multiple scales and outputs both category labels and bounding box coordinates.

Given the lack of texture features in terahertz images and the insufficient recognition accuracy caused by similar contours and mutual occlusion among different types of targets, an enhanced terahertz image target recognition algorithm, enhanced and occlusion-aware focus YOLOv5 (EOF-YOLOv5), based on YOLOv5 is proposed. The network architecture is illustrated in Figure 4. Firstly, an occlusion-aware attention mechanism(OCA) is incorporated prior to the cross stage partial network with 3 convolutions (C3) module in the original YOLOv5 neck network. This mechanism integrates two complementary branches: a channel attention branch, which models interdependencies across feature channels, and an occlusion-aware branch, which captures spatial

 doi:10.20944/preprints202509.1760.v1

occlusion patterns. This dual-branch design effectively suppresses background interference while enhancing feature responses in visible target regions. Secondly, the original complete intersection over union (CIoU) loss function is replaced with focal-efficient intersection over union (Focal-EIoU) to improve the measurement accuracy of similarity between predicted bounding boxes and ground truth bounding boxes. EOF-YOLOv5 is not a simple combination of technologies, but an organic whole: the EOF framework (comprising the OCA mechanism and Focal-EIoU Loss) ensures the model can "see" and focus on the effective parts of the target, while also guiding it to strive to "learn" the most challenging samples. These two components synergize across feature extraction spatial weighting and loss function optimization to collectively enhance the recognition of terahertz images.

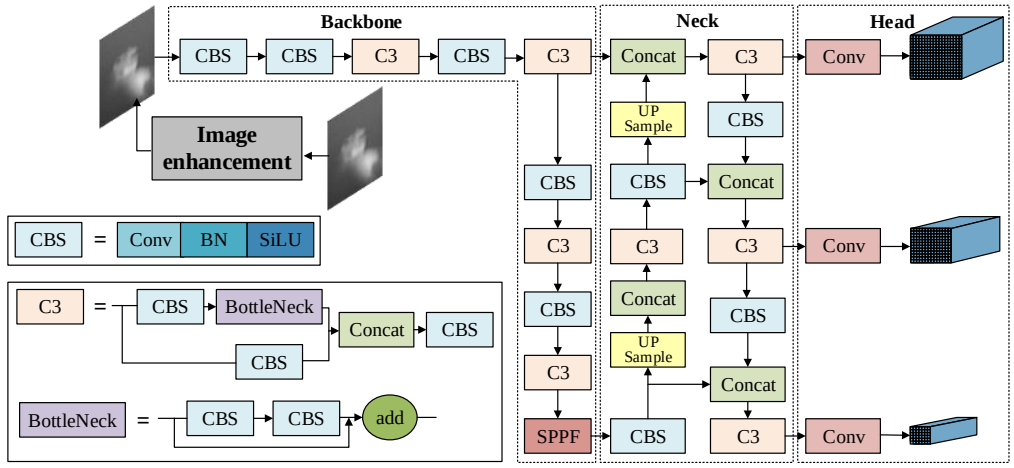


Figure 3. YOLOv5 network structure diagram.

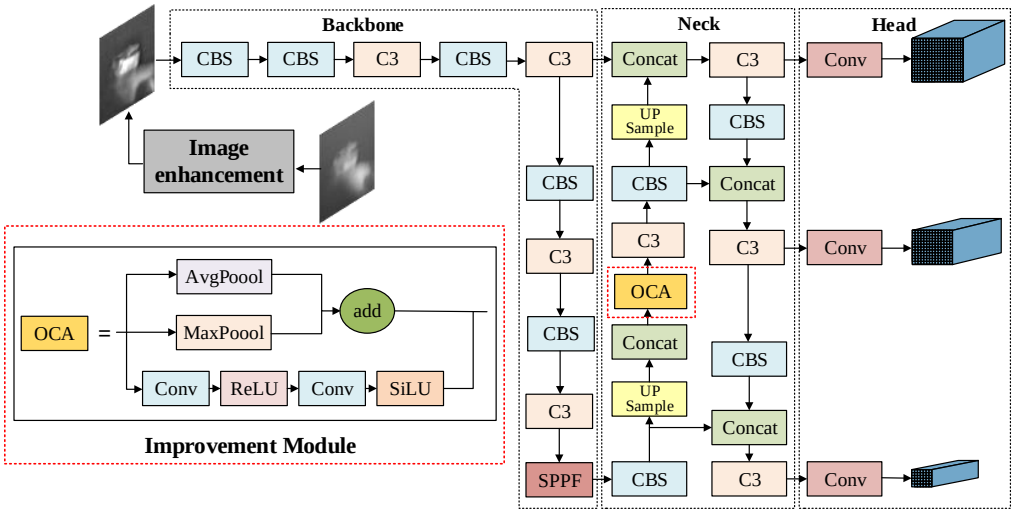


Figure 4. Structure of the EOF-YOLOv5 Model.

2.3.1. Occlusion-Aware Context Attention

Image enhancement preprocessing can significantly improve target contrast and clarity in terahertz images, thereby facilitating more efficient detection and recognition by the model. However, occlusion may result in partial coverage of the target by other objects, leading to the complete loss of the corresponding image information. The attention mechanism is a widely adopted technique in computer vision aimed at enhancing the ability of network models to focus on regions of interest within input images [33].

To improve the efficiency of attention mechanisms, this paper proposes an Occlusion-Aware Context Attention (OCA) mechanism, which assigns adaptive weight coefficients to individual channel. This method effectively suppresses background interference while amplifying feature

 doi:10.20944/preprints202509.1760.v1

responses from visible regions of the target [34,35]. The OCA module is integrated after the concatenation (Concat) operation in the Feature Pyramid Network (FPN) and before the C3, meaning that the input to the OCA module is the feature map enriched with multi-scale information after fusion by the FPN. Since OCA performs "point-by-point" weight adjustment, it enhances the feature representation capability of the output while keeping the feature map size unchanged, providing a seamless interface for the subsequent C3 network layer. The occlusion-aware attention mechanism, as illustrated in Figure 5, consists of two main components: a channel attention branch, which models interdependencies across channel, and an occlusion-aware branch, which captures spatially structured occlusion information. The channel attention branch utilizes both global average pooling (AvgPool) and global max-pooling (MaxPool) to compress spatial information [36]. Specifically, AvgPool is utilized to smooth details while preserving the overall distribution, whereas MaxPool is designed to capture local extrema of small-scale and occluded features. The integration of both pooling operations enables a multi-scale feature representation, thereby enhancing the perception and characterization of small-scale occluded features while mitigating information loss associated with single-pooling strategies. The inter-channel dependencies are modeled through two successive 1×1 convolutional layers incorporating ReLU activation functions, ultimately generating the channel attention weights $M_{ch}(X)$ via the Sigmoid function.

$$M_{ch}(X) = \sigma(W_2 Re LU(W_1(\frac{AvgPool(X) + MaxPool(X)}{2}))) \tag{9}$$

In the equation (9), $W_1 \in R^{C/r \times C}$, $W_2 \in R^{C \times C/r}$ represents the learnable weight, r denotes the reduction ratio, and σ signifies the Sigmoid activation function.

The Occlusion-Aware Branch directly applies a 1×1 convolution (with ReLU activation) to the input feature map X to extract spatial occlusion patterns, subsequently generating a pixel-wise mask $M_{occ}(X)$ through the Sigmoid function.

$$M_{occ}(X) = \sigma(W_4 \operatorname{Re} LU(W_3 X) \tag{10}$$

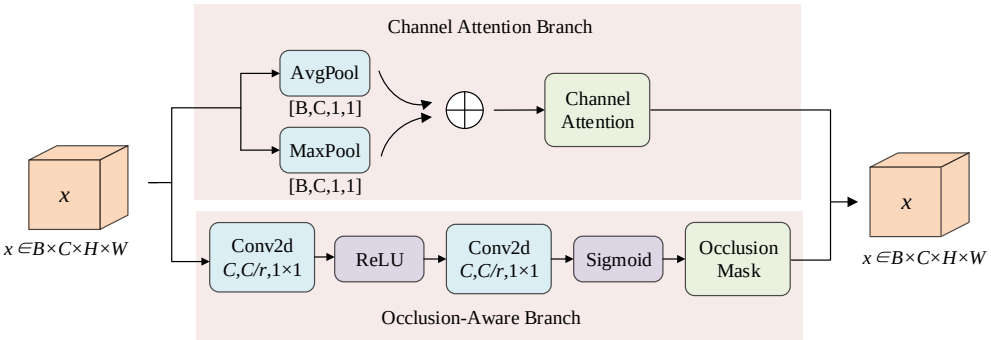
In the equation (10), $W_3 \in R^{C/r \times C}$, $W_4 \in R^{C \times C/r}$ represents the corresponding compression/expansion weighting factor.

The final output is a feature map with dual attention weighting:

$$Output = X \odot M_{occ}(X) \odot M_{ch}(X) \tag{11}$$

In the equation (11), $X \in R^{C \times H \times W}$ represents the input feature map, and $\odot$ denotes the element-wise multiplication operation.

The OCA module achieves feature calibration through multiplicative fusion by concurrently learning spatial occlusion relationships and inter-channel dependencies. Specifically, the occlusion branch maintains the original feature map resolution and captures local occlusion patterns using 1×1 convolutions, while the channel branch extracts channel statistics via global pooling, integrating both average and max-pooled features. The final output is refined under dual attention constraints, which amplify discriminative features and suppress occluded or noisy regions.



 doi:10.20944/preprints202509.1760.v1

Figure 5. OCA structural diagram.

2.3.2. Focal EIoU

In YOLOv5, the Complete Intersection over Union (CIoU) is employed as the loss function. However, since CIoU treats all samples equally, it may cause the gradients from easy samples to dominate the learning process, thereby drowning out the contribution of hard samples. [37]. The CIoU loss function is defined as shown in Formula (12):

$$Loss_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + av \tag{12}$$

$$IoU = \frac{A \cap B}{A \cup B} \tag{13}$$

In this context, A and B represent the predicted box and the ground truth box, respectively; IoU denotes the Intersection over Union (IoU) between the predicted box and the ground truth box; $\rho^2(b, b^{gt})$ signifies the Euclidean distance between the centroids of the predicted box and the ground truth box, with b being the centroid of the predicted box and the centroid of the ground truth box; c^2 represents the area of the smallest enclosing region that contains both the predicted box and the ground truth box; $\frac{\rho^2(b, b^{gt})}{c^2}$ addresses the optimization of issue IoU; a is utilized to balance the scale; and v serves as a parameter describing the consistency of the aspect ratio between the predicted box and the ground truth box.

In scenarios involving occluded targets, the CIoU bounding box loss function demonstrates limitations in accurately capturing positional relationships between objects. This often leads to imprecise target localization, along with an increased rate of false positives and missed detections. Furthermore, the slow convergence rate of the loss function adversely impacts the accuracy of localization. In this context, the Focal-EIoU loss function is introduced in this paper [38], which significantly enhances the measurement accuracy of similarity between predicted bounding boxes and ground truth bounding boxes. The schematic diagram of the Focal-EIoU parameters is presented in Figure 6.

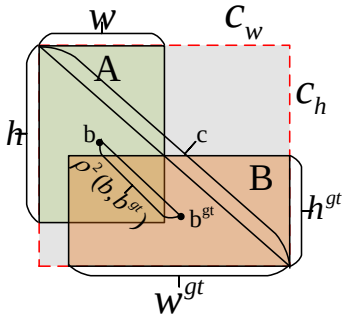


Figure 6. Schematic Diagram of Focal-EIoU Loss Function Parameters.

The EIoU metric facilitates the alignment of the aspect ratio between predicted and ground-truth bounding boxes [39], thereby accelerating the convergence rate of the loss function. The computational formula for the EIoU loss value is presented in the following equation:

$$\begin{aligned} Loss_{EIoU} &= Loss_{IoU} + Loss_{dis} + Loss_{asp} \\ &= 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{c_w^{\ 2}} + \frac{\rho^2(h, h^{gt})}{c_h^{\ 2}} \end{aligned} \tag{14}$$

 doi:10.20944/preprints202509.1760.v1

In the equation (14), c_w and c_h denote the width and height of the minimum bounding box encompassing the predicted box and the ground truth box, respectively; w and h represent the width and height of the predicted box; w^{gt} and h^{gt} indicate the width and height of the ground truth box; $Loss_{IoU}$, $Loss_{dis}$, and $Loss_{asp}$ correspond to the overlap loss, center distance loss, and side length loss between the ground truth box and the predicted box, respectively.

Focal loss incorporates an adjustment factor within the loss function that is correlated with sample difficulty. For easily classifiable samples, this factor remains relatively small, whereas for challenging samples, the adjustment factor is significantly larger. The mathematical expression for the Focal-EIoU bounding box loss function is as follows:

$$Loss_{Focal\_EIoU} = IoU^{\gamma} Loss_{EIoU} \tag{15}$$

In the equation (15), γ represents the hyperparameter utilized for balancing the weights of positive and negative samples.

Focal-EIoU optimizes the loss weighting mechanism to enhance the model's focus on challenging regression samples, such as small or occluded targets, while mitigating excessive attention to straightforward samples. This approach mitigates the issue of gradient dominance by easy samples during bounding box regression, thereby enhancing the model's ability to learn from hard samples.

2.4. Performance Evaluation Indicators

This paper uses standard deviation (*SD*) and average gradient (*AG*) as evaluation indicators for measuring texture detail and edge intensity information before and after image preprocessing [40].

$$SD = \frac{\frac{1}{N}\sqrt{\sum\left(F_i(x,y) - \sum F_i(x,y)/N\right)^2}}{\sum F_i(x,y)/N} \tag{16}$$

In the equation (16), N represents the number of pixels, and $F_i(x,y)$ denotes the grayscale value at coordinate (x,y).

$$AG = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{\frac{\nabla F_x^2(i,j) + \nabla F_y^2(i,j)}{2}} \tag{17}$$

$$\nabla F_x(i,j) = F(i,j) - F(i-1,j) \tag{18}$$

$$\nabla F_y(i,j) = F(i,j) - F(i,j-1) \tag{19}$$

In this context, M and N represent the pixel dimensions of the image, while ∇F_x and ∇F_y correspond to the gradients of the image in the horizontal and vertical directions, respectively.

The evaluation of model performance is conducted along two primary dimensions: accuracy and speed. For accuracy assessment, this study employs Precision (P), Recall (R), Average Precision (AP), and Mean Average Precision (mAP). Speed is evaluated in frames per second (FPS), while model efficiency is measured using parameters (Params) and floating point operations (FLOPs). The relevant calculation formulas are shown in Equations (20) to (24).

$$P = \frac{TP}{TP + FP} \tag{20}$$

$$R = \frac{TP}{TP + FN} \tag{21}$$

 doi:10.20944/preprints202509.1760.v1

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{22}$$

$$mAP = \frac{\sum_{i=1}^{C} AP_i}{C} \tag{23}$$

$$FPS = \frac{1}{time} \tag{24}$$

In the equation, TP denotes true positives, FP represents false positives, FN stands for false negatives, and time indicates the duration required to process a single image.

3. Experimental and Results Analysis

3.1. Experimental Environment and Parameter Settings

The experimental setup in this study was configured with the following specifications: NVIDIA Quadro RTX 4000 GPU, with CUDA versions 12.1 installed, running on the Windows 10 operating system. The Python version is 3.12, and the deep learning framework used is PyTorch 2.2.1. All original images were uniformly preprocessed to a 640×640 pixel input size. Specifically, while preserving the original aspect ratio, symmetric padding was applied to the shorter side of each image to achieve the target dimensions. The hyperparameters were configured as follows: the batch size is 16, the maximum learning rate is 0.01, the learning rate momentum is 0.98, and the training epochs are set to 100.

3.2. Image Pre-Processing Results

The results of the image enhancement preprocessing are presented in Figure 7, which shows that the target contours become more distinct and the contrast between the target and background is significantly improved after preprocessing. Table 5 provides a comparative analysis of the original images and the preprocessed images in terms of standard deviation and average gradient. As indicated in Table 3, the numerical values of the enhanced images show a marked improvement over the original images. These results demonstrate that the proposed image enhancement preprocessing methodology effectively improves both the visual quality and structural detail of the images.

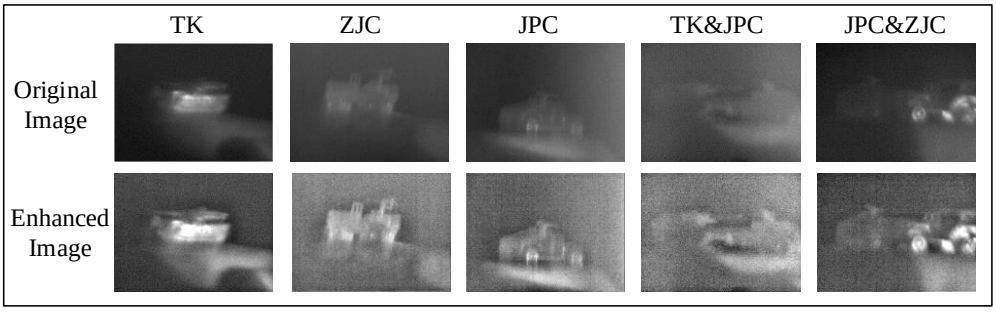


Figure 7. Comparative Analysis of Original Image and Preprocessed Image Enhancement Results.

Table 3. Evaluation Metrics for Image Preprocessing Results.

Metric	Image	TK	ZJC	JPC	TK&JPC	JPC&ZJC
SD	Original	27.09	15.81	26.81	21.16	25.71
	Enhanced	**41.19**	**35.07**	**41.67**	**30.21**	**35.26**
AG	Original	27.18	36.15	40.13	30.64	30.61
	Enhanced	**53.93**	**162.51**	**177.53**	**77.20**	**69.32**

 doi:10.20944/preprints202509.1760.v1

3.3. Ablation Experiment

To systematically evaluate the effectiveness of each enhancement module in the target recognition task on terahertz datasets using the improved algorithm, and to investigate the contribution of individual improvement strategies to overall performance, this section presents a comprehensive ablation study. First, maintain consistent parameters across all experiments. Train and test models using both the original images and the preprocessed images with image enhancement, to assess the impact of image enhancement preprocessing on model performance. Second, using enhanced images as the baseline dataset, we progressively introduce the improved OCA attention mechanism and Focal EIoU loss function. The analysis focuses on precision (P), recall (R), mean average precision at IoU threshold 0.5 (mAP50) and 0.5:0.95 (mAP50-95), frames per second (FPS), parameter count (Params), and floating point operations (FLOPs) obtained from the identical test set. The experimental results are summarized in Table 4, where the "√" symbol denotes the inclusion of the respective module. A total of four ablation experiments were conducted. The bolded numerical values in the table represent the optimal metrics for each evaluation criterion.

Table 4. Ablation experiment.

Yolov5	Image Enhancement	OCA	Focal EIoU	P (%)	R (%)	mAP50 (%)	mAP50-95 (%)	FPS	FLOPs (G)	Params (M)
√	-	-	-	59.6	66.2	59.2	34.5	89.28	**15.8**	**7.01**
√	√	-	-	66.7	73.1	75.5	48.2	85.47	**15.8**	**7.01**
√	√	√	-	72	73.3	77.7	49.9	84.74	15.9	7.08
√	√	√	√	**79.3**	**80.6**	**83.7**	**52.2**	**87.7**	15.9	7.08

Analysis of the ablation experiment results in Table 4 reveals that after applying enhancement preprocessing to the original images, the model's mAP50 improved by 16.3 percentage points from 59.2% to 75.5%. This demonstrates that image preprocessing significantly enhances model performance. After introducing the improved OCA attention mechanism module into the model, the mAP50 metric exhibited a further improvement of 2.2 percentage points, reaching 77.7%, albeit with a relatively modest enhancement. Upon the final integration of the Focal-EIoU loss function, the model achieved optimal performance, with mAP50 increasing to 83.7%, representing a significant enhancement of 11.8 percentage points compared to the baseline model. It is noteworthy that during the progressive integration of various enhancement modules, the computational cost of the model (FLOPs and Params) exhibited only a marginal increase. Remarkably, the final model achieved an FPS of 87.7, surpassing the performance observed with the sole inclusion of the image enhancement module (85.47). This demonstrates that these improvements not only enhance detection accuracy but also maintain superior computational efficiency. Comprehensive analysis indicates that the incorporation of image enhancement and various improvement modules has made positive contributions to the model's performance.

 doi:10.20944/preprints202509.1760.v1

3.4. Comparative Experiments

Comparative experiments were conducted using YOLOv5s, YOLOv8n, YOLOv10n, YOLOv12n, and the enhanced algorithm EOF-YOLOv5 on the same preprocessed dataset and under identical configurations. Based on the log files preserved during the training process on a dataset without occluded targets, comparative curves for mAP50, Precision, and Recall of the five models were plotted, as illustrated in Figure 8.

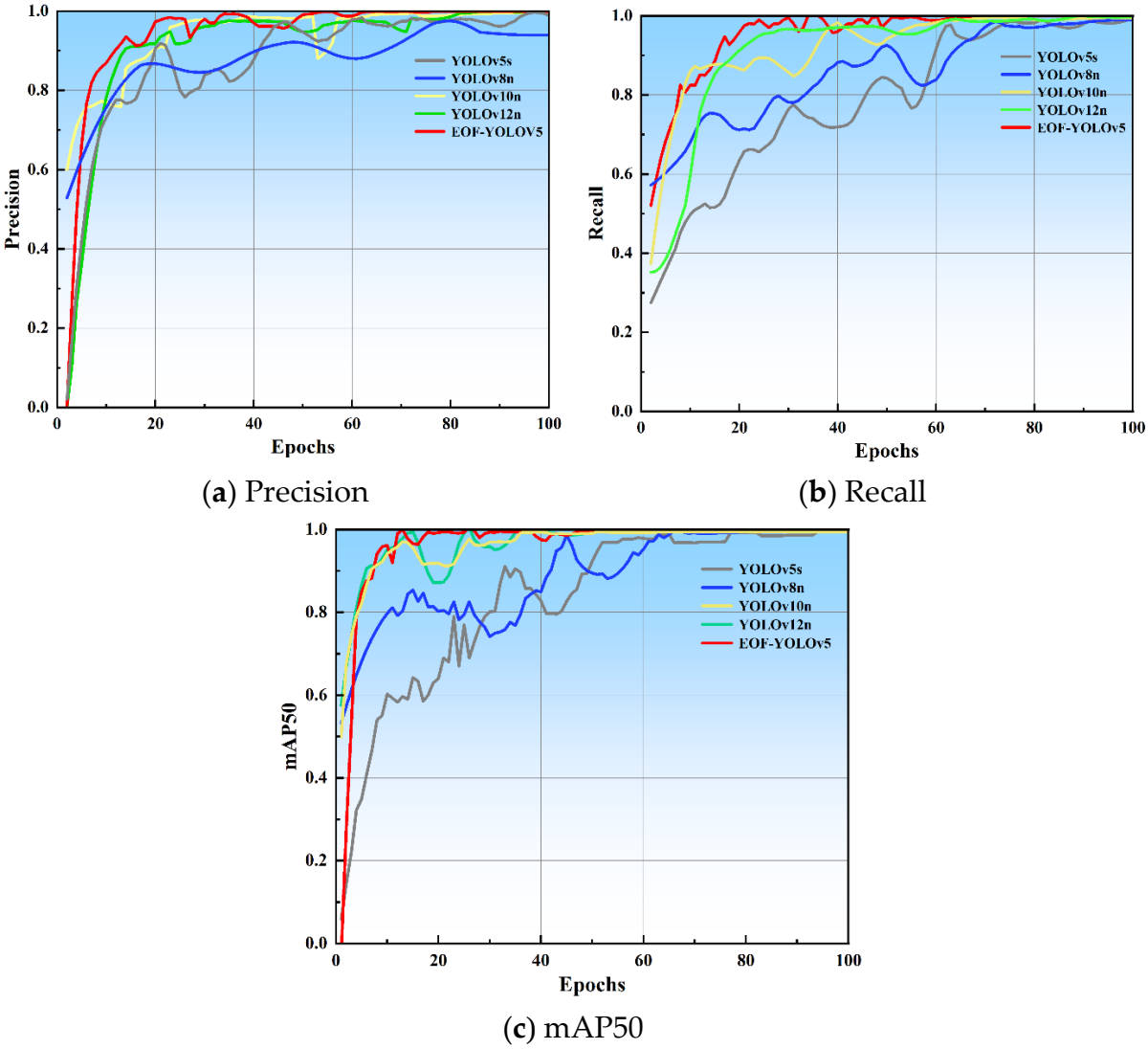


Figure 8. Training Result Curves of Various Comparison Algorithms.

As illustrated in Figure 8(a) by the precision curve, the enhanced algorithm (represented by the red curve) consistently maintains the highest precision throughout the training process, demonstrating rapid convergence, thereby exhibiting more stable and reliable classification capabilities. In contrast, the comparative algorithms exhibit lower precision levels during the initial training phase, indicating that the enhanced algorithm offers a distinct advantage in reducing false detection rates. As illustrated in Figure 8(b) 's recall curve, the enhanced algorithm (represented by the red curve) demonstrates a rapid increase in recall rate during the initial training phase, ultimately stabilizing at a significantly higher level. This indicates its superior capability in capturing more genuine targets and reducing missed detections. Figure 8(c) illustrates the mAP50 curve, which integrates the performance metrics of precision and recall, serving as a key indicator for evaluating the overall efficacy of the algorithm. It is evident that the enhanced algorithm (represented by the red curve) exhibits a significantly higher mAP50 value compared to all benchmark algorithms, coupled

 doi:10.20944/preprints202509.1760.v1

with a more rapid convergence rate, thereby demonstrating its superior accuracy and robustness in detection tasks.

The comparative experimental data obtained from the test set are shown in Table 5. where the bolded values indicate the optimal performance for each evaluation metric.

Table 5. Comparison of results with other methods.

Methods	P (%)	R (%)	mAP50 (%)	mAP50-95 (%)	FPS	FLOPs (G)	Params (M)
YOLOv5s	66.7	73.1	75.5	48.2	85.47	15.8	7.01
YOLOv8n	68.6	69.8	66.3	44.9	99.15	6.20	2.38
YOLOv10n	65.7	66.8	73.4	42.6	80.17	7.31	2.62
YOLOv12n	62.5	72.3	69.7	42.0	**102.80**	**5.74**	**2.33**
EOF-YOLOv5	**79.3**	**80.6**	**83.7**	**52.2**	87.7	15.9	7.08

As illustrated in Table 5, the enhanced YOLOv5 algorithm EOF-YOLOv5 proposed in this study achieves a 12.6% improvement in precision over the original YOLOv5s algorithm, indicating its efficacy in reducing false positives caused by background interference. Furthermore, the recall rate exhibits a 7.5% enhancement, substantiating the algorithm's superior capability in reducing missed detections, particularly excelling in scenarios involving occlusions and small-scale targets. In terms of comprehensive detection performance, the model demonstrates significant improvements with a 8.2% increase in mAP50 and a 4% enhancement in mAP50-95. These advancements not only elevate the detection accuracy under standard IoU thresholds but also refine the quality of bounding box regression under stringent localization requirements. Notably, these performance gains are achieved while maintaining nearly constant parameter counts (Params) and computational complexity (FLOPs). The proposed algorithm demonstrates significant improvements over the state-of-the-art lightweight object detection models, including YOLOv8n, YOLOv10n, and YOLOv12n. Specifically, it achieves precision enhancements of 10.7%, 13.6%, and 16.8%, respectively. The recall rates are improved by 10.8%, 13.8%, and 8.8%. Furthermore, the mAP50 metrics show increases of 17.4%, 10.3%, and 14%, while the mAP50-95 metrics are enhanced by 7.3%, 9.6%, and 10.2%. These results validate the superior performance of the improved algorithm in multi-scale object detection and precise localization. The FPS of the proposed method exhibits a reduction of 11.45 frames per second compared to YOLOv8n, while demonstrating a significant improvement of 7.53 frames per second over YOLOv10n and a decrease of 15.1 frames per second relative to YOLOv12n. YOLOv12n maintains the minimal parameter count (Params) and computational complexity (FLOPs) at 5.74G and 2.33M respectively. YOLOv8n and YOLOv10n demonstrate comparable parameter counts and computational complexities, while the modified algorithm exhibits approximately double the parameter count and computational complexity of both YOLOv8n and YOLOv10n. The analysis indicates that the proposed modified algorithm achieves substantial detection accuracy enhancement through a moderate increase in computational resource allocation.

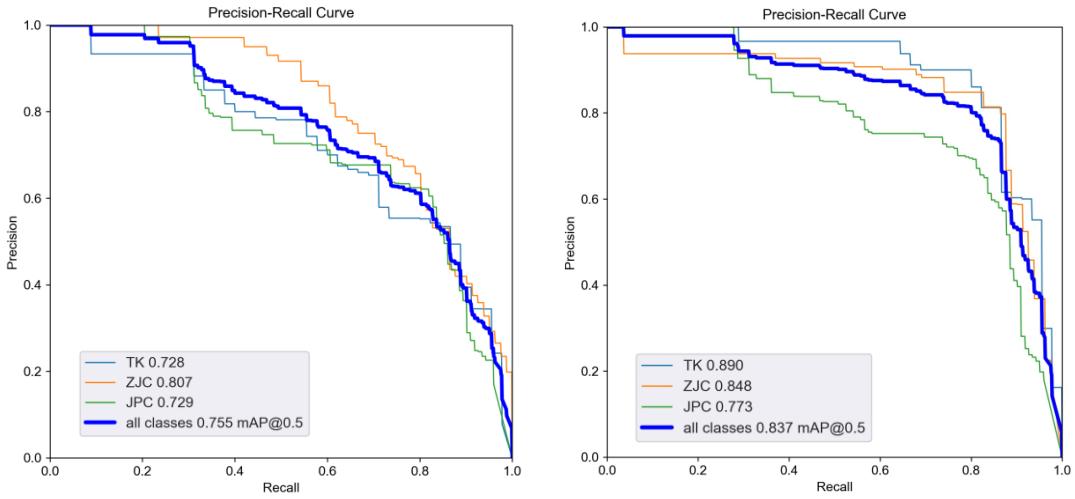
3.5. Visualization Results for Different Occlusion Conditions

The Precision-Recall (P-R) curve provides a visual representation of the trade-off between precision and recall across different threshold values for a model. An optimal P-R curve should be lie as close as possible to the upper-right corner, indicating that a larger area under the curve corresponds to higher precision and recall performance of the model. Figure 10 illustrates a comparative analysis of the Precision-Recall (P-R) curves between the original YOLOv5 model and the enhanced EOF-YOLOv5 model.

As shown in Figure 9, the improved model EOF-YOLOv5 significantly outperforms the original model across multiple key metrics: the average precision for tank targets increased from 0.728 to 0.89, for armored vehicles from 0.807 to 0.848, and for jeeps from 0.729 to 0.773. The mean average precision (mAP@0.5) across all categories increased from 0.755 to 0.837, corresponding to a relative

 doi:10.20944/preprints202509.1760.v1

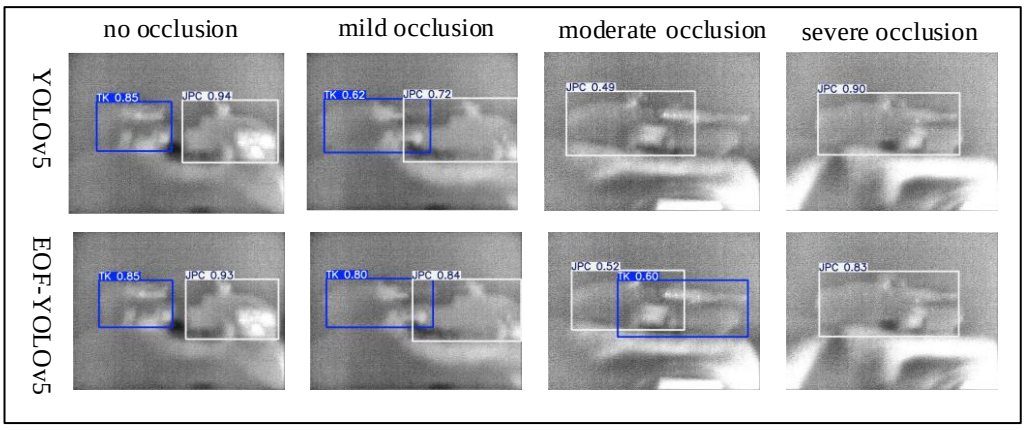
improvement of 8.2%. The results indicates that EOF-YOLOv5, through algorithmic enhancements, significantly improves detection accuracy while maintaining a high recall rate. This is reflected in a distinct upward and rightward shift of the P-R curve, demonstrating superior overall detection performance. These findings are supported by the ablation study results summarized in Table 4, collectively confirming the efficacy of the proposed improvements.



(**a**) YOLOv5 (**b**) EOF-YOLOv5

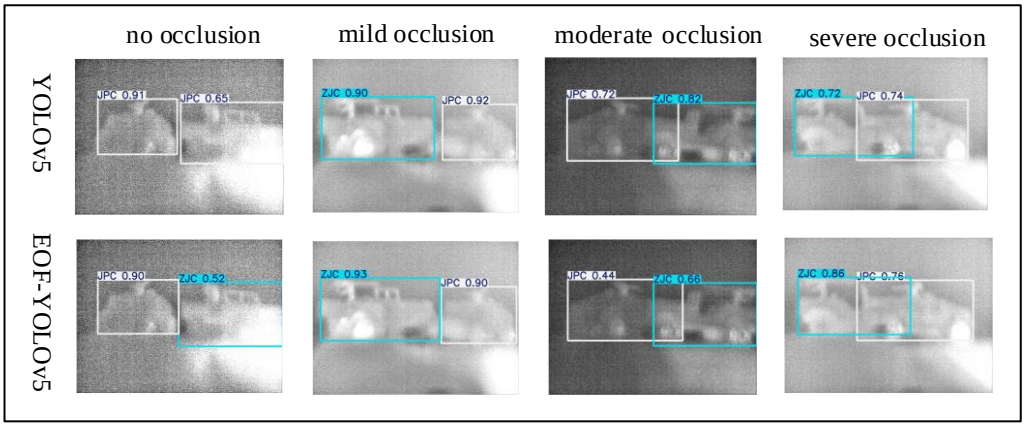
Figure 9. Comparative Analysis of Precision-Recall Curves between the Original Model and the Enhanced Model.

To facilitate a more intuitive evaluation of detection performance after model enhancement, a visual comparison was conducted between the proposed EOF-YOLOv5 model and the baseline YOLOv5 model using the same test dataset. To quantify the impact of target occlusion on model performance, occlusion levels were categorized into four classes based on ratio of the mutual occlusion area to the target's intrinsic area: no occlusion, mild occlusion (occlusion area < 30%), moderate occlusion (30% ≤ occlusion area ≤ 60%), and severe occlusion (occlusion area > 60%). Based on this classification criterion, the system performed a comprehensive analysis of feature robustness and detection accuracy across four typical occlusion scenarios (no occlusion, mild occlusion, moderate occlusion, and severe occlusion) to elucidate the correlation between occlusion severity and model performance. The comparative results of the actual detection performance between the two models are illustrated in Figures 10.



(**a**) JPC&TK

 doi:10.20944/preprints202509.1760.v1



(**b**) JPC&ZJC

Figure 10. Comparative Analysis of Detection Performance between the Original Model and the Enhanced Model.

Figure 10(a) presents a comparative analysis of the Jeep (JPC) and Tank (TK) detection performance under four occlusion scenarios. Both models correctly identify objects under no occlusion and mild occlusion conditions, with the EOF-YOLOv5 model yielding higher confidence scores than the YOLOv5 model. Under moderate occlusion, the YOLOv5 model fails to detect the Tank, whereas the EOF-YOLOv5 model maintains accurate detection under both mild and moderate occlusion conditions. Under severe occlusion, the extensive loss of salient visual features due to large occluded areas considerably weakens the mapping relationship between the residual local features and the actual object categories, leading to a significant decline in detection performance. Figure 10(b) presents a comparative analysis of the detection results for the Jeep (JPC) and Armored Vehicle (ZJC) across the four occlusion scenarios. In cases of no occlusion, the YOLOv5 model misclassifies armored vehicles as jeeps due to high morphological similarity in the overall contour features between the targets. The EOF-YOLOv5 model, by enhancing image contours and strengthening the representation capability of high-frequency contour information in images, demonstrates superior detection accuracy.

4. Conclusions

This paper proposes an enhanced YOLOv5 algorithm, designated EOF-YOLOv5, for target recognition in terahertz images. Focusing on military targets including tanks, armored vehicles, and jeep models, we constructed a comprehensive terahertz dataset by acquiring and annotating a large number of terahertz images using a phased-array terahertz near-field imaging system. To address the challenges of blurred edge contours and low target contrast in terahertz images, a stationary wavelet transform enhancement algorithm was implemented to preprocess the original dataset. To mitigate the reduction in effective feature information and the decline in recognition accuracy caused by inter-object occlusion, an occlusion-aware attention mechanism was incorporated prior to the C3 module within the YOLOv5 backbone network. This enhancement strengthens the feature response in visible regions and improves recognition accuracy. Furthermore, in scenarios where targets are occluded, the CIoU bounding box loss function struggles to distinguish the positions between targets, leading to inaccurate target localization. We replaced the CIoU loss function with the Focal-EIoU loss function to mitigate excessive focus on simple samples and alleviate the problem of simple samples dominating the gradient in bounding box regression, thereby enhancing the model's learning capability for challenging samples. The EOF-YOLOv5 algorithm achieves mean average precision (mAP) of 83.7%, a recall of 80.6%, and a precision of 79.3%, demonstrating a significant improvement in recognition performance over the baseline YOLOv5 model. Furthermore, comparative experiments with other state-of-the-art algorithms confirm the superiority of the proposed method. The results demonstrate that the presented algorithm achieves outstanding recognition accuracy in

 doi:10.20944/preprints202509.1760.v1

terahertz near-field radar image target identification tasks. Its core mechanism and processing strategy are particularly well-suited to the distinctive properties of terahertz images.

The proposed EOF-YOLOv5 algorithm, with its enhanced capabilities in accurately detecting occluded objects and improving recognition precision in low-contrast scenarios], demonstrates significant potential for application in ground-to-ground reconnaissance systems based on terahertz near-field imaging. Despite achieving solid performance, EOF-YOLOv5 still has some limitations. First, the computational complexity of the current model remains challenging for certain extremely resource-constrained embedded platforms, such as mobile robots and drones. Future work will explore more advanced model pruning, quantization, or neural architecture search techniques to further reduce model size and computational requirements, facilitating practical deployment. Second, research will focus on unsupervised or semi-supervised domain adaptation methods, enabling the model to rapidly adapt to new, unlabeled target scenarios.

Author Contributions: Conceptualization, J.C. and Z.Y.; methodology, J.C. and W.W.; software, L.G.; validation, L.Y. and L.W.; formal analysis, Y.L.; investigation, Z.Y.; data curation, J.C.; writing—original draft preparation, J.C.; writing—review and editing, Z.Y.; visualization, J.C.; supervision, Z.Y.; project administration, W.W.; funding acquisition, Z.Y and Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported in part by the Key Scientific Research Plan of Education Department of Shaanxi [23JY035], The Youth Innovation Team of Shaanxi Universities [K20220184], and Natural Science Foundation of Shaanxi Province [2025JC-YBMS-744].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: Author Yan Feng was employed by the company North Electronic-Optics Company Limited. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Sun, L.; Zhao, L.; Peng, R.Y. Research progress in the effects of terahertz waves on biomacromolecules. *Mil. Med. Res.* **2021**, 8, 28.
2. Fleming, J.W. High-Resolution Submillimeter-Wave Fourier-Transform Spectrometry of Gases. *IEEE Trans. Microw. Theory Tech.* **1974**, 22, 1023–1025.
3. Siegel, P.H. Terahertz technology. *IEEE Trans. Microw. Theory Tech.* **2002**, 50, 910–928.
4. Horiuchi, N. Terahertz technology: Endless applications. *Nature Photonics* **2010**, 4, 140. https://doi.org/10.1038/nphoton.2010.16
5. Ran, Z.; Yuanmeng, Z.; Cunlin, Z. Target aided identification in passive human THz-image. *High Power Laser Part. Beams* **2014**, 26, 132–136.
6. Xin, Z.; Yuanmeng, Z.; Chao, D.; Cunlin, Z. Study on the passive terahertz image target detection. *Acta Opt. Sin*. **2013**, 33, 83–88.
7. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, 39, 1137–1149.
8. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
9. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.***2017.** 2999–007.

10. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779-788.
11. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
12. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv: 1804. 02767.
13. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv: 2004. 10934.
14. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475.
15. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J. Yolov10: Real-time end-to-end object detection. *Adv. Neural Inf. Process. Syst.* **2024**, 37, 107984–108011.
16. Xiao, H.; Zhang, R.; Wang, H.; Zhu, F.; Zhang, C.; Dai, H.; Zhou, Y. R-PCNN method to rapidly detect objects on THz images in human body security checks. In Proceedings of the 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation, Guangzhou, China, 8–12 October 2018.
17. Yang, X.; Wu, T.; Zhang, L.; Yang, D.; Wang, N.; Song, B.; Gao, X. CNN with spatio-temporal information for fast suspicious object detection and recognition in THz security images. *Signal Process.* **2019**, 160, 202–214.
18. Kovbasa, M.; Golenkov, A.; Sizov, F. Neural Network Application to the Postal Terahertz Scanner for Automated Detection of Concealed Items. In Proceedings of the 2020 IEEE Ukrainian Microwave Week (UkrMW), Kharkiv, Ukraine, 21-25 September 2020.
19. Xu, F.; Huang, X.; Wu, Q.; Zhang, X.; Shang, Z.; Zhang, Y. YOLO-MSFG: Toward real-time detection of concealed objects in passive terahertz images. *IEEE Sens. J.* **2021**, 22, 520–534.
20. Danso, S.A.; Liping, S.; Hu, D.; Afoakwa, S.; Badzongoly, E.L.; Odoom, J.; Muhammad, O.; Mushtaq, M.U.; Qayoom, A.; Zhou, W. An optimal defect recognition security-based terahertz low resolution image system using deep learning network. *Egypt. Inform. J.* **2023**, 24, 100384.
21. Yu, H.; Yang, Q.; Wang, H. Recognition method for space target components in terahertz radar images based on improved YOLOv5. *Proc. SPIE* **2024**, 13495, 1349504.
22. Ge, Z.; Zhang, Y.; Jiang, Y.; Ge, H.; Wu, X.; Jia, Z.; Wang, H.; Jia, K. Lightweight YOLOv7 Algorithm for Multi-Object Recognition on Contrabands in Terahertz Images. *Applied Sciences* **2024**, 14, 1398.
23. Zeng, Z.; Wu, H.; Chen, M.; Luo, S.; He, C. Concealed hazardous object detection for terahertz images with cross-feature fusion transformer. *Opt. Laser Eng.* **2024**, 182, 108454.
24. Yang, L.; Wang, H.; Zeng, Y.; Liu, W.; Wang, R.; Deng, B. Detection of Parabolic Antennas in Satellite Inverse Synthetic Aperture Radar Images Using Component Prior and Improved-YOLOv8 Network in Terahertz Regime. *Remote Sens.* **2025**, *17*, 604.
25. Cheng, A.; Wu, S.; Liu, X.; Lu, H. Enhancing concealed object detection in active THz security images with adaptation-YOLO. Sci. Rep. **2025**, 15, 2735.
26. Kumar, A.; Tomar, H.; Mehla, V.K.; Komaragiri, R.; Kumar, M. Stationary wavelet transform based ECG signal denoising method. *ISA Trans.* **2021**, 114, 251–262.
27. Fan, X.; Ding, W.; Li, X.; Li, T.; Hu, B.; Shi, Y. An Improved U-Net Infrared Small Target Detection Algorithm Based on Multi-Scale Feature Decomposition and Fusion and Attention Mechanism. *Sensors* **2024**, 24, 4227.
28. Liu, Y.; Jiang, H.; Liu, C.; Yang, W.; Sun, W. Data-augmented wavelet capsule generative adversarial network for rolling bearing fault diagnosis. *Knowl.-Based Syst.* **2022**, 252, 109439.
29. Liu, H.; Xu, Z.; Wei, Y.; Han, K.; Peng, X. Multispectral non-line-of-sight imaging via deep fusion photography. *Sci. China Inf. Sci.* **2025**, 68, 1–19.
30. Liu, J.; Zhou, X.; Wan, Z.; Yang, X.; He, W.; He, R.; Lin, Y. Multi-Scale FPGA-Based Infrared Image Enhancement by Using RGF and CLAHE. *Sensors* **2023**, 23, 8101.

 doi:10.20944/preprints202509.1760.v1

31. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-Time End-to-End Object Detection. *arXiv* **2024**, abs/2405.14458.
32. Tian, Y.; Ye, Q.; Doermann, D. Yolov12: Attention-centric real-time object detectors. *arXiv preprint* **2025**, arXiv:2502.12524.
33. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, 452, 48–62.
34. Li, Y.; Liu, Y.; Zhang, H.; Zhao, C.; Wei, Z.; Miao, D. Occlusion-Aware Transformer With Second-Order Attention for Person Re-Identification. *IEEE Trans. Image Process.* **2024**, 33, 3200–3211.
35. Su, Y.; Sun, R.; Shu, X.; Zhang, Y.; Wu, Q. Occlusion-aware detection and re-id calibrated network for multi-object tracking. *arXiv preprint* **2023**, arXiv:2308.15795.
36. Liu, B.; Ge, R.; Zhu, Y.; Zhang, B.; Zhang, X.; Bao, Y. IDAF: Iterative Dual-Scale Attentional Fusion Network for Automatic Modulation Recognition. *Sensors* **2023**, 23, 8134.
37. Huang, P.; Tian, S.; Su, Y.; Tan, W.; Dong, Y.; Xu, W. IA-CIOU: An Improved IOU Bounding Box Loss Function for SAR Ship Target Detection Methods. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, 17, 10569–10582.
38. Zhang, Y.F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *arXiv preprint* **2021**, arXiv:2101.08158.
39. Yu, Q.; Han, Y.; Han, Y.; Gao, X.; Zheng, L. Enhancing YOLOv5 Performance for Small-Scale Corrosion Detection in Coastal Environments Using IoU-Based Loss Functions. *J. Mar. Sci. Eng.* **2024**, *12*, 2295.
40. Sun, Y.; Zhao, Z.; Jiang, D.; Tong, X.; Tao, B.; Jiang, G.; Kong, J.; Yun, J.; Liu, Y.; Liu, X.; Zhao, G.; Fang, Z. Low-Illumination Image Enhancement Algorithm Based on Improved Multi-Scale Retinex and ABC Algorithm Optimization. *Front. Bioeng. Biotechnol.* **2022**, 10, 865820.