remote sensing

MDPI

Article

SAR Small Ship Detection Based on Enhanced YOLO Network

Tianyue Guan [1,2], Sheng Chang [1,*], Chunle Wang [1] and Xiaoxue Jia [1]

[1] Space Microwave Remote Sensing System Department, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China; guantianyue20@mails.ucas.ac.cn (T.G.); jiaxx@aircas.ac.cn (X.J.)
[2] School of Electronics, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: changsheng19@mails.ucas.ac.cn

Abstract: Ships are important targets for marine surveillance in both military and civilian domains. Since the rise of deep learning, ship detection in synthetic aperture radar (SAR) images has achieved significant progress. However, the variability in ship size and resolution, especially the widespread presence of numerous small-sized ships, continues to pose challenges for effective ship detection in SAR images. To address the challenges posed by small ship targets, we propose an enhanced YOLO network to improve the detection accuracy of small targets. Firstly, we propose a Shuffle Re-parameterization (SR) module as a replacement for the C2f module in the original YOLOv8 network. The SR module employs re-parameterized convolution along with channel shuffle operations to improve feature extraction capabilities. Secondly, we employ the space-to-depth (SPD) module to perform down-sampling operations within the backbone network, thereby reducing the information loss associated with pooling operations. Thirdly, we incorporate a Hybrid Attention (HA) module into the neck network to enhance the feature representation of small ship targets while mitigating the interference caused by surrounding sea clutter and speckle noise. Finally, we add the shape-NWD loss to the regression loss, which emphasizes the shape and scale of the bounding box and mitigates the sensitivity of Intersection over Union (IoU) to positional deviations in small ship targets. Extensive experiments were carried out on three publicly available datasets—namely, LS-SSDD, HRSID, and iVision-MRSSD—to demonstrate the effectiveness and reliability of the proposed method. In the small ship dataset LS-SSDD, the proposed method exhibits a notable improvement in average precision at an IoU threshold of 0.5 (AP50), surpassing the baseline network by over 4%, and achieving an AP50 of 77.2%. In the HRSID and iVision-MRSSD datasets, AP50 reaches 91% and 95%, respectively. Additionally, the average precision for small targets (AP) exhibits an increase of approximately 2% across both datasets. Furthermore, the proposed method demonstrates outstanding performance in comparison experiments across all three datasets, outperforming existing state-of-the-art target detection methods. The experimental results offer compelling evidence supporting the superior performance and practical applicability of the proposed method in SAR small ship detection.

Keywords: synthetic aperture radar (SAR); small ship detection; you only look once (YOLO); re-parameterized convolution

check for updates

Academic Editors: Zhenwei Shi, Kefeng Ji, Mingjin Zhang, Xiangguang Leng, Haohao Ren and Zhongzhen Sun

Received: 3 January 2025
Revised: 15 February 2025
Accepted: 25 February 2025
Published: 27 February 2025

Citation: Guan, T.; Chang, S.; Wang, C.; Jia, X. SAR Small Ship Detection Based on Enhanced YOLO Network. *Remote Sens.* **2025**, *17*, 839. https://doi.org/10.3390/rs17050839

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Synthetic aperture radar (SAR) is an active microwave imaging sensor that offers significant advantages over optical sensors. Firstly, SAR operates independently of weather conditions and ambient illumination, enabling continuous all-day and all-weather observations of the Earth. Additionally, SAR possesses the capability to detect concealed targets

and gather critical information in complex environments. Due to these benefits, SAR has become an indispensable information acquisition platform in remote sensing [1–4]. Its applications span various domains, including topographic mapping, agricultural monitoring, disaster detection, and ocean monitoring [5–7]. Notably, the utilization of SAR imagery for ship detection has been applied across various fields, including marine surveillance, fishery management, disaster rescue, and others [8].

Traditional SAR ship detection methods mainly encompass constant false alarm rate (CFAR)-based methods [9], visual saliency-based methods [10], global threshold-based methods [11], polarimetry decomposition-based methods [12], and wavelet transform-based methods [13]. These traditional methods heavily rely on manually crafted features and limited shallow learning representations, rendering them highly vulnerable to variations in background statistics. In complex scenarios, such as coastal areas characterized by small islands or rocky outcrops, traditional SAR ship detection algorithms struggle to identify effective features necessary for achieving optimal results and exhibit restricted generalization capabilities.

With the rapid development of deep learning technology [14], researchers have increasingly integrated this technology into the study of ship detection tasks in SAR images. Initially, convolutional neural networks (CNNs) were employed to enhance the detection performance of traditional SAR image ship detection methods, such as constant false alarm rate (CFAR) [15,16]. Subsequently, as more SAR ship detection training datasets became available, researchers shifted their focus towards end-to-end SAR image ship detection methods that are entirely based on CNNs. These methods demonstrated exceptionally high levels of efficiency and accuracy in detection. Ship detection methods based on CNNs can be categorized into two types: one is a two-stage detection network exemplified by Region-CNN (R-CNN) [17], while the other is a one-stage detection network represented by You Only Look Once (YOLO) and its subsequent iterations [18–22]. The two-stage detection network employs selective search techniques to generate candidate regions for further analysis. In contrast, the one-stage detection network eliminates the region proposal generation phase, thereby simplifying the object detection problem into a regression problem, resulting in faster detection speeds compared to two-stage detectors.

Due to the development of CNN-based object detection methods, numerous advanced SAR ship detection methods have been proposed in recent years. Cui et al. [23] introduced a novel CenterNet that incorporates a spatial shuffle-group enhanced attention module, enabling the extraction of more refined semantic features while effectively suppressing noise to mitigate false alarms caused by coastal and inland interferences. Shan et al. [24] proposed a multi-layer deep dense network that includes an improved DenseNet block along with a SimAM attention mechanism, which adeptly adapts to the specific speckle noise distribution present in SAR images and dynamically enhances spatial features, thereby allowing the entire network to exhibit robust generalization capabilities and high accuracy. Hu et al. [25] presented an anchor-free framework utilizing a balanced attention network aimed at enhancing ship detection across various scales. This framework incorporates both local and non-local attention modules to achieve an equilibrium between local and non-local feature representations throughout the entire network's operation. With ongoing optimizations and widespread applications of YOLO-series algorithms, an increasing number of researchers are employing YOLO algorithms to optimize ship detection methods. Sun et al. [26] proposed a novel YOLO-based SAR ship detector using bi-directional feature fusion and angular classification (BiFA-YOLO). The novel bi-directional feature fusion module (Bi-DFFM) can efficiently aggregate multi-scale features through bi-directional (top-down and bottom-up) information interaction, which is helpful for detecting multi-scale ships. Zhao et al. [27] integrated a fused attention C2fSE module along with a DenseASPP

module into the baseline YOLOv8n architecture to bolster both feature extraction and fusion capabilities. Luo et al. [28] introduced SHIP-YOLO, a lightweight SAR ship detection model that is built upon the YOLOv8n framework. The model incorporates GhostConv, a lightweight convolutional layer designed to replace conventional convolutions, and integrates the re-parameterized RepGhost bottleneck structure within the C2f module. This innovative approach effectively reduces both the parameter count and computational complexity of the model. Additionally, they implemented WIoU and attention mechanisms to enhance detection accuracy. Dai et al. [29] proposed LHSDNet, a lightweight yet high-accuracy SAR ship image detection network. This architecture employs GhostHGNetV2 as its feature extraction backbone and includes a lightweight feature fusion module to minimize overall computational load. Furthermore, parameter sharing is utilized in the feature extraction component, while the design of the detection head prioritizes lightness to conserve computing resources further.

SAR images often contain a significant number of small targets, which can lead to a considerable decline in detection accuracy. Nevertheless, the existing methods for identifying small targets exhibit limited detection capabilities. The challenges associated with detecting small ships in maritime contexts have garnered considerable attention from researchers. Some scholars have validated the detectability of small boats within SAR images through studies focused on specific types of vessels. Lanz et al. [30,31] prepared test vessels, including thin inflated rubber ships and simulated fully occupied refugee vessels, in an effort to identify the most effective detectors for various satellite platforms. The study presented in [32] highlights the advantages of utilizing very-high-resolution TerraSAR-X data acquired in staring spotlight mode for the detection of small boats. Shin et al. [33] proposed a method to improve the detection capability of small ships through the integration of multiple polarization images and the application of an adaptive threshold technique. In the domain of Convolutional Neural Networks (CNNs), researchers have conducted numerous explorations. Wang et al. [34] redesigned the feature extraction network and proposed a path augmentation fusion network aimed at integrating spatial and semantic information across different scales through both bottom-up and top-down approaches. Yang et al. [35] introduced an anchor-free enhanced FCOS method that employs a multi-level feature attention mechanism along with a feature refinement and reuse module to extract effective features while refining those related to small ships, thereby facilitating the effective detection of small ships near shorelines. Zhou et al. [36] integrated the transformer self-attention mechanism into their backbone network, utilizing channel attention and spatial attention to enhance feature space integration. Hu et al. [37] proposed a transformer-based dynamic sparse attention module and a small target-friendly detection head to improve the focus and extraction of small ship features.

Research on CNN-based ship detection has made progress in both accuracy and efficiency, and researchers have begun to focus on difficulties in ship detection such as small-sized ship detection, but there still exist several intractable challenges that require solutions. First, ships exhibit rich diversity in real SAR images, and variations in scale continue to pose challenges for detection. In particular, small ships occupy a limited number of pixels within SAR images, making them susceptible to being overlooked, which can lead to missed detections. Second, due to the intrinsic imaging mechanism of SAR, speckle noise is an unavoidable characteristic in SAR images. In complex backgrounds, small targets are more susceptible to being submerged by sea clutter and speckle noise, causing a decline in detection accuracy. Third, there is an inherent contradiction between detection accuracy and detection speed. To effectively identify small targets, it is necessary to increase the complexity of the model, which poses challenges in meeting the requirements for real-time ship detection.

Based on the analysis presented above, this paper proposes an enhanced YOLO network with the Shuffle Re-parameterization (SR) module, space-to-depth (SPD) module, Hybrid Attention (HA) module, and shape-NWD loss. The novel network replaces the C2f module with the SR module, which facilitates the propagation of target information throughout the feature hierarchy in both backbone and neck networks. The SPD module is employed to perform down-sampling operations to retain more available information. The incorporation of the HA module contributes to an overall enhancement of ship-related features. And the shape-NWD loss proves advantageous for detecting small-sized ships. The main contributions of this article can be summarized as follows:

1. We propose a Shuffle Re-parameterization block (SRB) that integrates channel shuffle with a re-parameterized convolution block (RepConv). This approach enhances the capabilities of feature extraction during the training phase, while concurrently reducing memory consumption and inference time in the inference phase. To further improve the detection capabilities of the model, we substitute the original C2f module with the Shuffle Re-parameterization module (SR) constructed using the proposed RepConv and SRB.
2. We employ a Space-to-Depth module (SPD) for down-sampling operations to mitigate information loss associated with small targets, thereby enhancing the detection accuracy of small-sized ships.
3. An attention module called Hybrid Attention module (HA) is proposed, which leverages both frequency-domain and spatial-domain information. This module enhances the focus on ship-related features while effectively minimizing the influence of irrelevant background interference.
4. We add the shape-NWD loss to enhance detection accuracy further, since it exhibits insensitivity to targets of varying scales, and is particularly well suited for measuring the similarity between small targets.

The subsequent sections of this paper are structured as follows. Section 2 elaborates on the proposed methodology; Section 3 presents the experimental outcomes; Section 4 discusses the results; and Section 5 is the conclusion.

2. Methodology

2.1. The Overall Network Architecture

YOLOv8 is a well-established and widely adopted one-stage anchor-free detection model that aims to achieve an optimal balance between accuracy and speed. It has demonstrated exceptional performance across various object detection tasks, including SAR ship detection. In comparison to anchor-based detection model YOLOv5, the anchor-free design of YOLOv8 is more adept at accommodating the multi-scale characteristics of ship targets. Meanwhile, while YOLOv11 demonstrates superior performance in general object detection tasks, YOLOv8 exhibited greater flexibility in complex scenes and achieved enhanced results in ship detection within SAR images. Given the computational complexity and efficiency requirements associated with SAR ship detection, we selected YOLOv8n as the baseline model for constructing our network. The overall architecture of the network is illustrated in Figure 1, which consists of three components: a backbone network for feature extraction, a neck for multi-scale feature fusion, and three decoupled prediction heads dedicated to classification and regression. The detailed network structure of the reserved YOLOv8 module, specifically the SPPF module and detection heads, is also illustrated in Figure 1. The modules that deviate from the baseline network are highlighted in red and will be discussed in greater detail in the following sections.

First, we introduce a re-parameterized convolution block (RepConv) to propose the Shuffle Re-parameterization module (SR). The SR module replaces the C2f module in

YOLOv8, aiming to enhance the feature representation capabilities of the detection network and extract texture features more effectively. Second, we incorporate the Space-to-Depth module (SPD) to replace the down-sampling convolution layer in the backbone network, intending to reduce the information loss associated with small targets and preserve more valuable features. The output of the backbone network is subsequently fed to the neck network, which comprises a top-down path FPN and a bottom-up Path Aggregation Network (PAN), to extract contextual information. We propose a Hybrid Attention (HA) module, strategically positioned between FPN and PAN. The HA module is designed to enhance features related to ships and suppress surrounding interference by implementing frequency-domain and spatial-domain attention mechanisms. After the neck network, the decoupled head employs an anchor-free structure for category and position prediction. This approach aims to mitigate the inaccuracies in predictions arising from the multi-scale characteristics of ships. Finally, we incorporate the shape-NWD loss into the bounding box regression loss. The shape-NWD loss integrates the concept of shape-IoU with the Normalized Gaussian Wasserstein Distance (NWD) to enhance detection performance for small targets. The proposed method is applied for small ship detection in SAR imagery, resulting in improved performance in detection accuracy.

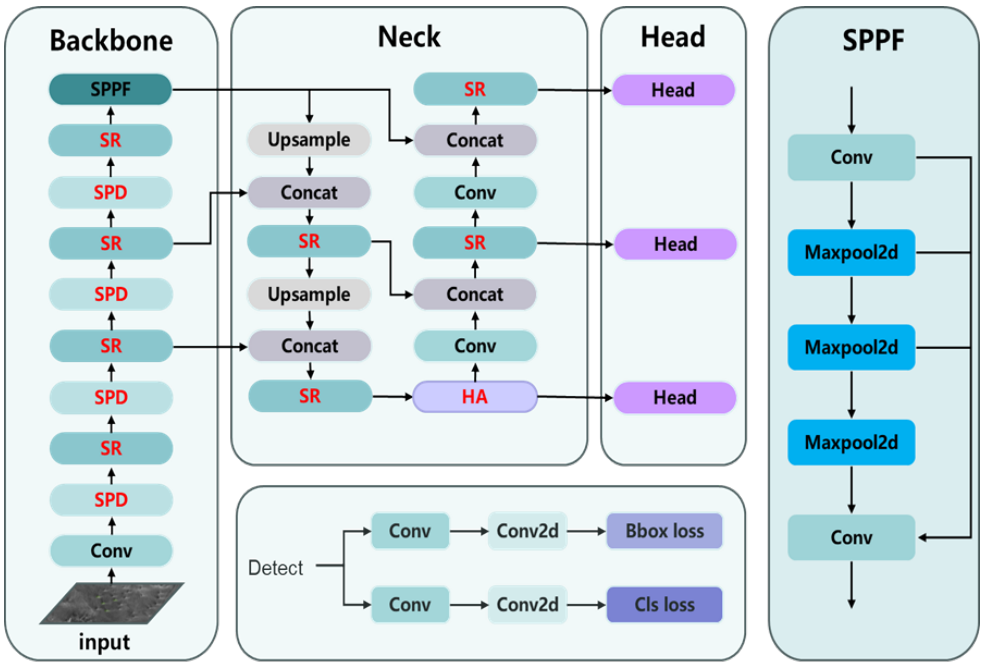


Figure 1. The overall architecture of the proposed method.

The rest of Section 2 is organized as follows. Section 2.2 introduces the SR module, including the RepConv block, the SRB, and the SR module. The SPD module is presented in Section 2.3. In Section 2.4, the HA module is described in detail. Section 2.5 introduces the shape-NWD loss.

2.2. The Shuffle Re-Parameterization Module

In Convolutional Neural Networks, deepening the depth of the network can improve the feature representation capability. However, as the depth of the network increases, there is a corresponding risk of losing feature information related to the target. Unlike optical images, the distinctive features of ship targets in SAR images are often less pronounced, which complicates the process of effective feature extraction. To enhance the feature representation capabilities and extract ship-related features more effectively, we propose the

re-parameterized convolution block (RepConv). This innovation is subsequently employed to construct the SRB and the SR module.

It is well established that integrating branches of varying scales and complexities can enrich the feature space, thereby enhancing the representational capacity of individual convolutions. In this way, the capability for feature representation can be improved without necessitating an increase in the depth of the network. We adhere to the six equivalent transformations proposed in DBB [38] to construct a novel re-parameterized convolution block (RepConv), as illustrated in Figure 2. The RepConv block comprises three branches: the 1×1 convolution branch, the sequential $1 \times 1 - 3 \times 3$ convolution branch, and the 1×1 convolution-average pooling branch. In the sequential $1 \times 1 - 3 \times 3$ convolution branch, we set the internal channels equal to the input. A Batch Normalization (BN) layer is applied after each convolutional or average pooling layer to introduce nonlinearity during the training phase. The RepConv block employs various transformations, including BN, average pooling, sequences of convolutions, and branch addition. This multi-branch topology architecture exhibits diverse receptive fields and varying complexities, facilitating the acquisition of rich feature information during the training phase. In the inference stage, the RepConv block is converted into a 3×3 convolution through structural re-parameterization to enable rapid inference.

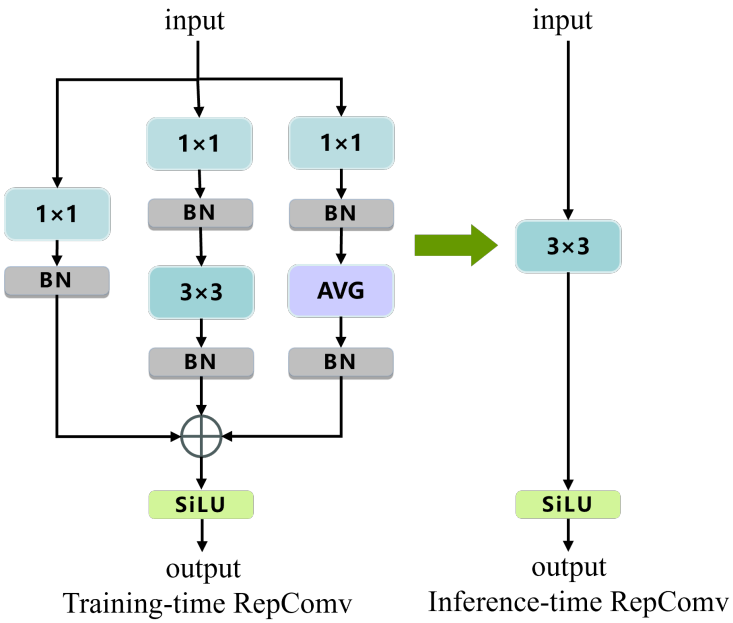
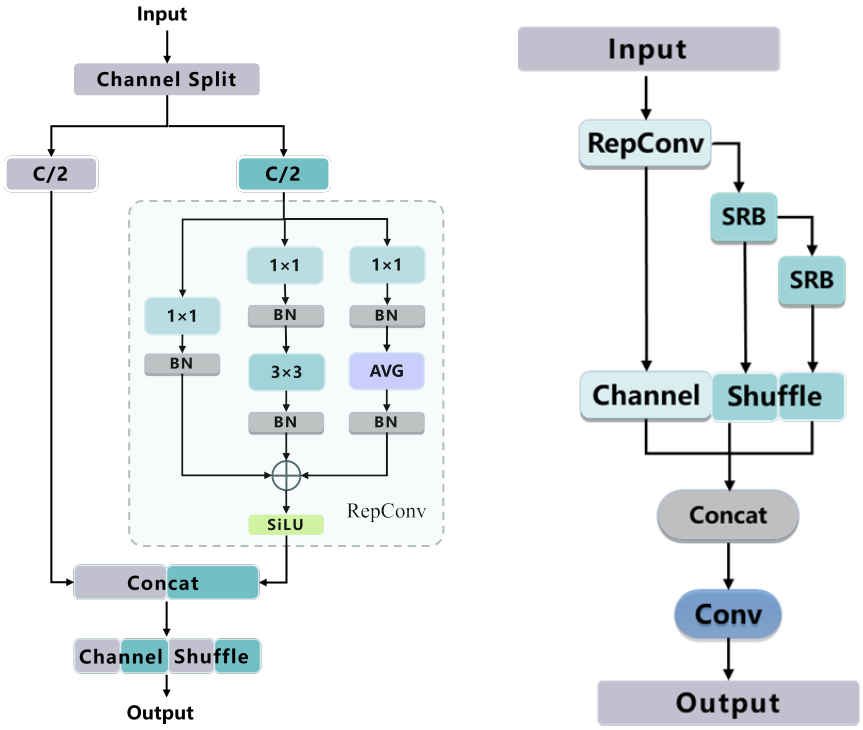


Figure 2. The structure of re-parameterized convolution block (RepConv).

Inspired by ShuffleNet [39], we construct a Shuffle Re-parameterization block (SRB) that integrates the RepConv block and the channel shuffle operation, as shown in Figure 3a. The input tensor is partitioned into two distinct channel-wise tensors with equal dimensions following the channel split operator. After conducting multi-branch training on one tensor, it is subsequently concatenated with the other tensor in a channel-wise manner. The channel shuffle mechanism operates on stacked group convolutions to enhance information fusion and facilitate informative feature representation.

The Shuffle Re-parameterization module (SR) is constructed by stacking RepConv and Shuffle Re-parameterization block (SRB), as illustrated in Figure 3b. Drawing inspiration from the One-Shot Aggregation (OSA) [40] module, we develop the SR module by aggregating features with varying receptive fields only once at the final feature maps. The SR module enhances the network's sensitivity to targets of different scales through

channel shuffling and feature stacking, thereby improving the model's detection capability for multi-scale targets. Due to the incorporation of the RepConv block and the channel shuffle operation, the SR module can extract abundant target features. We substituted the C2f module in the YOLOv8 network with the SR module. This modification enhances the feature extraction capabilities of the detection network, thereby improving its modeling proficiency for detection targets.



(a) SRB (b) SR

Figure 3. (**a**) The structure of Shuffle Re-parameterization block (SRB). (**b**) The structure of Shuffle Re-parameterization module (SR).

2.3. The Space-to-Depth Module

In SAR images, most ship targets appear as small entities with limited pixel size and restricted information content. Consequently, ship targets are prone to experiencing feature loss in the backbone network. Specifically, the down-sampling operation by convolution and pooling contributes significantly to information degradation and insufficient feature learning, which complicates the detection of small targets. To address this issue, we employ the SPD module [41] instead of the down-sampling convolution layer. The SPD module is composed of a 3×3 convolution layer, an SPD layer, and a 1×1 convolution layer, as shown in Figure 4.

The SPD layer partitions the input feature map into four distinct feature blocks, reducing both the length and width of the original input feature map by half while simultaneously increasing the number of channels by a factor of four. It plays a crucial role in reducing each spatial dimension of the input feature map while simultaneously enhancing the channel dimension, all without compromising the information contained within each channel. This is achieved by mapping features from the input feature map directly to channels. Consequently, as the spatial dimensions decrease, the size of the channel dimension increases. This transition layer effectively augments the depth of the feature map while preserving information, thereby mitigating issues related to information loss that are often encountered with traditional down-sampling layers. Following the SPD layer, a 1×1 convolution is applied as a standard convolution operation, which does not reduce the spatial dimensions

of the feature map. Thus, it facilitates further processing of features through learnable parameters while retaining more fine-grained details. This combination endows the SPD module with enhanced performance when addressing low-resolution images and small targets. By minimizing information loss associated with down-sampling, the SPD module significantly enhances detection accuracy for small targets.

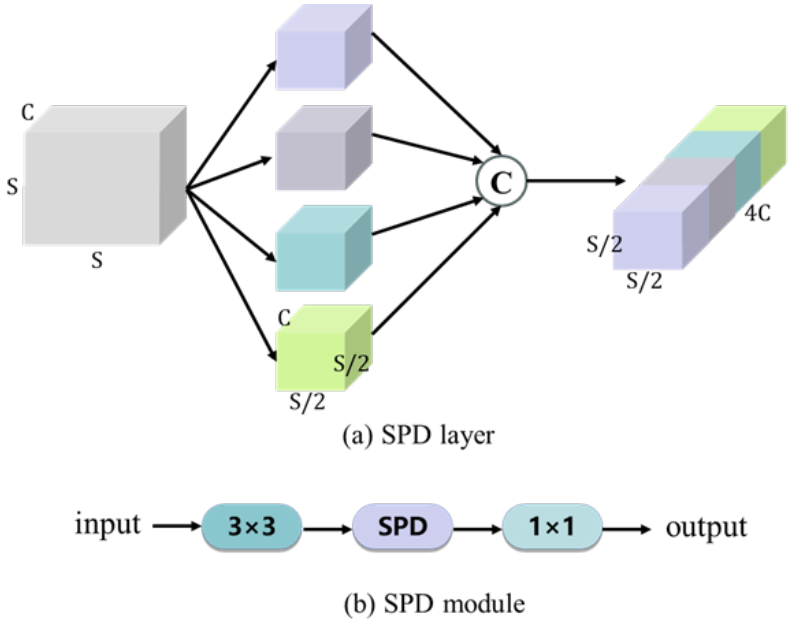


Figure 4. The structure of Space-to-Depth module. (**a**) is the structure of SPD layer. The character C represents the feature map's channel numbers, and the character S denotes the width and height of the input feature map. After the SPD layer, both the height and width of the feature map are reduced by half, while the number of channels is increased fourfold. (**b**) is the SPD module, which consists of the SPD layer. (SPD in the structure diagram (**b**) refers to SPD layer).

2.4. The Hybrid Attention Module

The object detection algorithm based on deep learning relies on the neck network for effective feature fusion. In YOLOv8, the PAN-FPN architecture generates feature maps that contain rich contextual and semantic information through bi-directional path fusion and horizontal connections. However, the long propagation path of PAN-FPN may result in the loss of ship texture details, which is detrimental to the detection of small ships. Furthermore, the widespread presence of clutter and noise heightens the probability that small ship targets may become entangled with surrounding interference or even be submerged, which presents additional challenges for achieving accurate feature representation. Consequently, we propose a Hybrid Attention (HA) module strategically positioned between the top-down and bottom-up pathways. This design aims to enhance the effective features of ship targets while mitigating information loss associated with long propagation paths.

Attention mechanisms have been widely used in computer vision tasks and have demonstrated remarkable effectiveness in producing discernible feature representations. In addition to channel attention mechanisms and spatial attention mechanisms, the frequency-domain information is proven to help the model fully understand the context. The Multi-axis External Weights module (MEW) [42] is a 2D-DFT-based attention mechanism that effectively extracts and fuses features to obtain more comprehensive global and local information. Figure 5a indicates that the feature map is divided into four branches in accordance with the channel dimension. Regarding the first three branches, features are converted to the frequency domain by means of 2D-DFT along three axes (Height–Width,

Channel–Width, and Channel–Height). Subsequently, the corresponding learnable weights are employed to multiply the frequency domain maps, and the map is transformed back to the spatial domain through the application of 2D-IDFT. For the fourth branch, DW convolution is used to acquire local information, and the feature map is concatenated along the channel dimension to restore the same size as the input. Eventually, the residual connection of the input is adopted to obtain the output. Given a feature map $X \in \mathbb{R}^{C\times H\times W}$ as input, the MEW mechanism can be expressed by the equations below.

$$x_1,\, x_2,\, x_3,\, x_4 = \mathrm{Split}(X) \tag{1}$$

$$x_1 = W_{(H,W)} \odot \mathrm{DFT}_{(H,W)}\left[x_1\right] \tag{2}$$

$$x_2 = W_{(C,W)} \odot \mathrm{DFT}_{(C,W)}\left[x_2\right] \tag{3}$$

$$x_3 = W_{(C,H)} \odot \mathrm{DFT}_{(C,H)}\left[x_3\right] \tag{4}$$

$$x_1' = \mathrm{IDFT}_{(H,W)}\left[x_1\right] \tag{5}$$

$$x_2' = \mathrm{IDFT}_{(C,W)}\left[x_2\right] \tag{6}$$

$$x_3' = \mathrm{IDFT}_{(C,H)}\left[x_3\right] \tag{7}$$

$$x_4' = \mathrm{DW}\left(x_4\right) \tag{8}$$

$$\mathrm{MEW}(X) = \mathrm{Concat}\left(x_1', x_2', x_3', x_4'\right) + X \tag{9}$$

where $\odot$ is the element-wise product. Split and Concat represent the split and concatenation operation along the channel dimension.

The Spatial Attention module (SA) [43] is illustrated in Figure 5b. SA is designed to emphasize or suppress different features within the feature map, enabling the network to focus on significant regions of the image. The spatial feature information is extracted by global maximum pooling and global average pooling, while convolution operations are employed to generate spatial attention weights. Given an input feature map $X \in \mathbb{R}^{C\times H\times W}$, the SA can be expressed by the equation below.

$$\begin{aligned}\mathrm{SA}(X) &= \mathrm{Conv}^{1\times 1}\Big(\mathrm{Conv}^{3\times 3}\big(\mathrm{Concat}\big(\mathrm{GMP}_{1d}(X), \mathrm{GAP}_{1d}(X)\big)\big)\Big) \\ &= \mathrm{Conv}^{1\times 1}\Big(\mathrm{Conv}^{3\times 3}\big(\mathrm{Concat}\big(x_s^{\mathrm{Max}}, x_s^{\mathrm{Avg}}\big)\big)\Big)\end{aligned} \tag{10}$$

where Concat represents the concatenation operation. GMP_{1d} and GAP_{1d} are 1D global maximum pooling operation and 1D global average pooling operation, respectively. $\mathrm{Conv}^{3\times 3}$ includes a 3×3 convolution operation and a SiLU activation function. $\mathrm{Conv}^{1\times 1}$ denotes the 2D conventional convolution with a kernel size of 1.

Combining SA and MEW, we introduce the hybrid attention (HA) module. The HA module employs MEW to transform feature maps into the frequency domain, thereby facilitating the acquisition of global information. The incorporation of frequency domain information addresses the limitation of CNNs in acquiring global information. Concurrently, SA is utilized to emphasize target-related features, thereby enhancing feature representa-

tion capabilities. The detailed architecture of the HA module is illustrated in Figure 5c. HA sequentially infers a frequency attention map and a spatial attention map, subsequently performing element-wise multiplication of the output features obtained from MEW and SA. The process can be expressed as follows:

$$F' = \mathrm{MEW}(F) \tag{11}$$

$$F'' = \mathrm{Resize}\big(\mathrm{SA}(\mathrm{F}')\big) \tag{12}$$

$$\mathrm{Output} = F' \odot F'' \tag{13}$$

where $\odot$ denotes element-wise multiplication. Resize represents resizing the size of the attention map to $\mathrm{H} \times \mathrm{W} \times \mathrm{C}$.

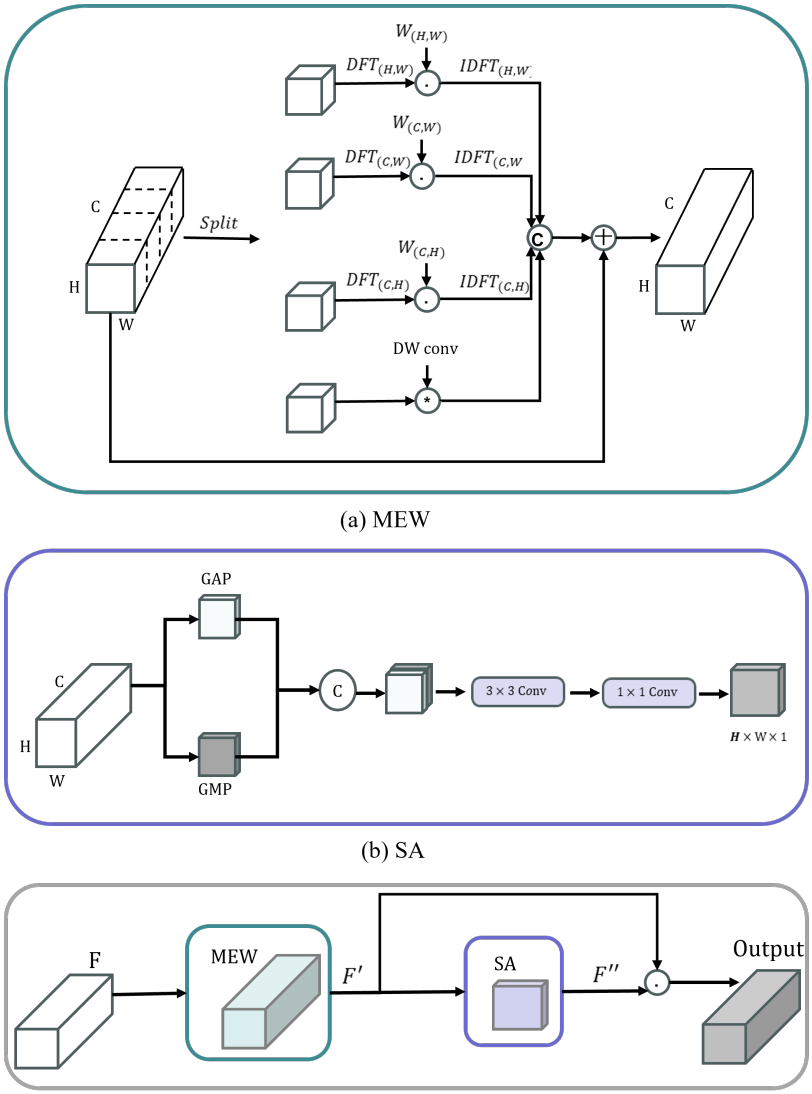


Figure 5. The structure of Hybrid Attention (HA). (**a**) is the Multi-axis External Weights module (MEW). (**b**) is the Spatial Attention module (SA). (**c**) is the Hybrid Attention (HA) module, which consists of the MEW module and the SA module.

2.5. The Loss Function

The loss functions used in the original YOLOv8 model include classification loss and regression loss. We concentrate on regression loss, which is composed of distribution focal loss (DFL) and Complete-IoU (CIoU) loss. The CIoU loss adds the aspect ratio criterion to the IoU to better fit the target frame on three geometric parameters: overlapping area, center point distance, and aspect ratio. The definition is as follows:

$$L_{\mathrm{CIoU}} = 1 - \mathrm{IoU} + \frac{\rho^2\left(b, b^{\mathrm{gt}}\right)}{c^2} + \alpha\, v \tag{14}$$

$$v = \frac{4}{\pi^2}\left(\arctan\left(\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}}\right) - \arctan\left(\frac{w}{h}\right)\right)^2 \tag{15}$$

$$\alpha = \frac{v}{(1 - \mathrm{IoU}) + v} \tag{16}$$

where $\rho^2\left(b, b^{\mathrm{gt}}\right)$ is the distance between the center points of the predicted and ground-truth bounding boxes, c is the diagonal distance of the minimum rectangular box containing two bounding boxes, α is a balance factor that increases with IoU.

Zhang et al. [44] pointed out that both the shape and scale factors of the bounding box regression sample influence the regression results. Notably, the small-scale bounding box was sensitive to variations in IoU values, and deviations along the short edge direction of the bounding box correspond to more pronounced changes in IoU values. Shape-IoU determines the loss by concentrating on the shape and scale of the bounding box itself, thus making the bounding box regression more precise. Given the large number of small-sized ship targets in SAR images, we use shape-NWD loss instead of CIoU loss. The shape-NWD integrates shape-IoU into NWD to avoid the sensitivity of IoU to the position deviation of small targets and improve detection performance. The shape-NWD loss is defined as follows:

$$L_{\mathrm{NWD^{shape}}} = 1 - \mathrm{NWD^{shape}} \tag{17}$$

$$\mathrm{NWD^{shape}} = e^{-\frac{D}{C}} \tag{18}$$

$$D = \sqrt{hh\left(x_c - x_c^{\mathrm{gt}}\right)^2 + ww\left(y_c - y_c^{\mathrm{gt}}\right)^2 + B} \tag{19}$$

$$B = \frac{(w - w_{\mathrm{gt}})^2 + (h - h_{\mathrm{gt}})^2}{weight^2}, \quad weight = 2 \tag{20}$$

$$hh = \frac{2 \times \left(h_{\mathrm{gt}}\right)^{\mathrm{scale}}}{\left(w_{\mathrm{gt}}\right)^{\mathrm{scale}} + \left(h_{\mathrm{gt}}\right)^{\mathrm{scale}}} \tag{21}$$

$$ww = \frac{2 \times \left(w_{\mathrm{gt}}\right)^{\mathrm{scale}}}{\left(w_{\mathrm{gt}}\right)^{\mathrm{scale}} + \left(h_{\mathrm{gt}}\right)^{\mathrm{scale}}} \tag{22}$$

where C is the constant associated with the dataset; scale is the scale factor, which is related to the scale of the target in the dataset; ww and hh are the weight coefficients in the horizontal and vertical directions, respectively, whose values are related to the shape of the gt box.

3. Experiments

3.1. Datasets and Experimental Settings

For object detection utilizing deep learning techniques, we adhere to the definition of object size established in the COCO dataset [45], focusing on the pixel proportion of the real object in the entire image, rather than their physical size. In this study, "small ship" refers specifically to those smaller pixel representations that occupy minor pixels in the whole image, rather than the real-world measurements of an actual ship (i.e., its length and width). Considering an image size of 800 × 800 pixels, targets with rectangular bounding box areas of less than 2342 pixels are classified as small objects, which constitute approximately 0.37% of the total 640,000 pixels.

The three publicly available remote sensing datasets used in this study are the LS-SSDD [46] dataset, the HRSID [47] dataset, and the iVision-MRSSD [48] dataset. The detailed information of the three datasets is shown in Table 1. HRSID and iVision-MRSSD only provided a single optimal resolution, while resolution in LS-SSDD refers to R. × A., where R. represents the range and A. denotes the azimuth.

Table 1. Detailed descriptions of LS-SSDD, HRSID, and iVision-MRSSD.

Dataset	LS-SSDD	HRSID	iVision-MRSSD
Date	2020	2020	2023
Image number	9000	5604	11,590
Ship number	6015	16,951	27,885
Image size	800 × 800	800 × 800	512 × 512
Number of sensors	1	3	6
Resolution(m)	5 × 20	0.5–3	0.5–100
Frequency bands covered	C	C, X	C, L, X

The LS-SSDD dataset is a large-scale background small ship detection dataset taken by Sentinel-1. It contains 15 large-scale SAR images of size 24,000 × 16,000 pixels and is divided into 9000 sub-images of size 800 × 800 pixels; the 9000 images are further divided into a training set and a test set in a 2:1 ratio. The large-scene small ship detection dataset meets the practical migration application of ship detection in large-scene space-borne SAR images in engineering.

The HRSID is a High-Resolution SAR image dataset sourced from Sentinel-1B, TerraSAR-X, and TanDEMX satellites. A total of 136 panoramic SAR images are cropped to 800 × 800 pixel sub-images under the overlapped ratio of 25%. The 5604 images are divided into a training set (65% SAR images) and a test set (35% SAR images) with the format of MS COCO.

The iVision-MRSSD dataset comprises 11,590 image tiles of size 512 × 512 pixels containing 27,885 ship examples. The dataset is produced by employing images from six distinct satellite sensors that cover a broad spectrum of the electromagnetic range, including C, L, and X band radar imaging frequencies. All of these sensors possess varying resolutions and imaging patterns. The dataset is randomly divided into training, validation, and test sets at a ratio of 70:20:10. The diverse circumstances allow the dataset to bring about a comprehensive understanding of the ship detection task in SAR satellite images.

In the ablation studies, SGD served as the optimizer with an initial learning rate of 0.01. LS-SSDD was trained for 160 epochs, while the other two datasets were trained

for 120 epochs. A batch size of 16 and an input size of 800 × 800 were used. For the comparative experiments, the training duration was set to 60 epochs, with a learning rate of 0.01, a weight decay of 0.0001, and a momentum of 0.9. The batch size and input size remained consistent with the ablation studies. All experiments were executed on a machine equipped with an RTX3090Ti (24G) GPU and an Intel Core i7-12400 CPU, running Windows 10, Python 3.11, and Cuda 11.7. The deep learning models were constructed using PyTorch 1.13. Additionally, comparisons were made using the MMDetection-3.1.0 framework [49] and the Ultralytics framework.

3.2. Evaluation Metrics

To quantitatively evaluate the performance of the proposed method, evaluation metrics such as precision (P), recall (R), average precision (AP), AP50, and average precision for different target sizes, that is, APs (Average Precision for Small objects), APm (Average Precision for Medium-sized objects), and APl (Average Precision for Large objects), were employed to assess the accuracy performance of the models.

Precision measures the ratio of accurately detected ship samples to all predicted ship samples, whereas recall quantifies the ratio of accurately detected ship samples to all annotated ship samples. The formulas for both are as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{23}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{24}$$

where TP (true positive) represents the count of correctly detected positive samples. FP (false positive) represents the count of incorrectly detected positive samples. FN (false negative) represents the count of incorrectly detected negative samples.

Average precision (AP) delineates the area under the precision–recall curve, composed of precision, recall, and the coordinate axis. The formula for AP is as follows:

$$\text{AP} = \int_0^1 P(R)\,\mathrm{d}R \tag{25}$$

where P represents precision, while R represents recall. AP50 corresponds to the average precision for an IoU threshold of 0.5; AP denotes the average precision when considering IoU ranges from 0.5 to 0.95 with a step of 0.05.

3.3. Ablation Experiments

To assess the performance of each module within the proposed method, we incrementally incorporated these modules into the baseline network. The baseline network utilized in this study is YOLOv8n, which adopts the same training strategy as that of the proposed method. We conducted ablation experiments on the LS-SSDD, HRSID, and iVision-MRSSD datasets to evaluate the effectiveness of the SR module, SPD module, HA module, and the enhanced loss function. For specific details, please refer to Tables 2–4.

Based on the results presented in Table 2, it is evident that on the LS-SSDD dataset, the proposed method achieves a 4.1% improvement in AP50 and a 1.7% improvement in AP compared to the baseline network. The incorporation of any one of the proposed modules individually results in an increase of 1.3% to 1.9% in AP50, along with an approximate rise of nearly 1% in AP compared to the baseline network. This demonstrates that each module plays a significant role in enhancing the performance of the proposed detection network. Figure 6 presents the precise–recall curves obtained through the incremental integration of the proposed modules. This representation more intuitively highlights the characteristics of each module and demonstrates the overall effectiveness of the proposed method.

Table 2. Results of ablation experiment on the LS-SSDD dataset.

Method	SR	SPD	HA	Loss	P	R	AP50	AP
Baseline	—	—	—	—	82.6	63.4	73.1	27.8
	✓	—	—	—	83.6	66.0	74.4	28.6
	—	✓	—	—	84.0	65.5	74.6	28.6
	—	—	✓	—	82.0	65.1	74.3	28.6
	—	—	—	✓	**84.7**	65.2	75.0	28.8
	✓	✓	—	—	83.8	66.1	75.0	29.0
	✓	✓	✓	—	83.7	68.5	76.2	29.1
Proposed	✓	✓	✓	✓	84.4	**69.1**	**77.2**	**29.5**

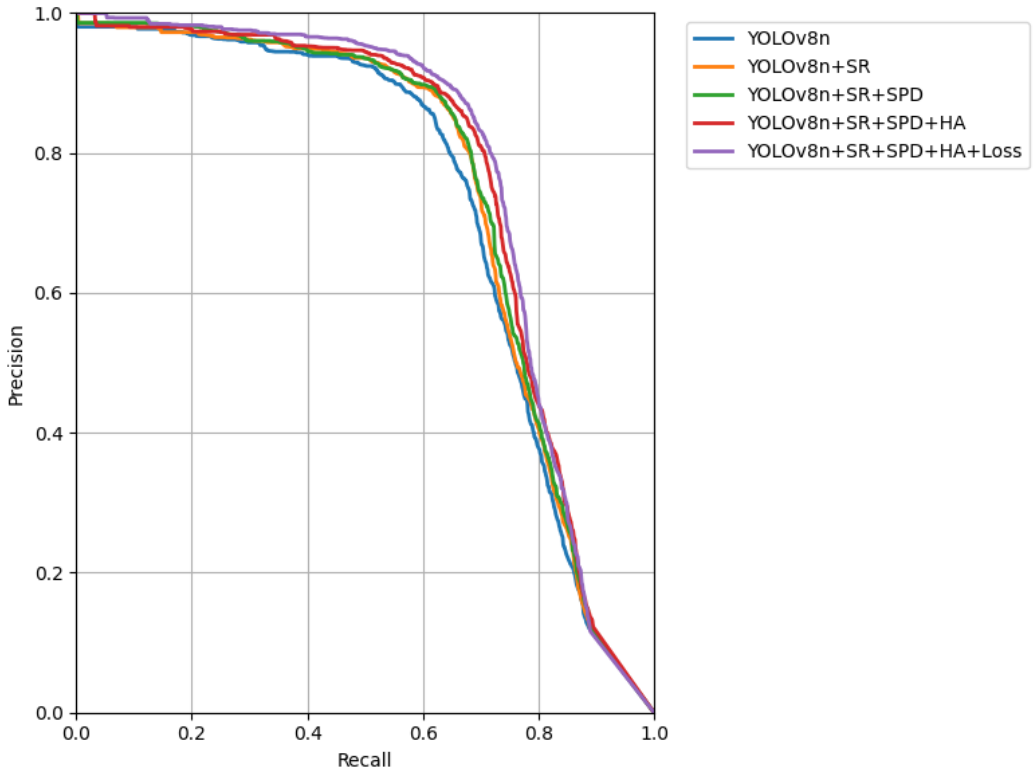


Figure 6. Precision–recall curves when adding different modules.

Figure 7 illustrates the visualized results of both the baseline and the proposed method across various scenarios involving small targets on LS-SSDD. Specifically, the first scenario is characterized by noisy far-sea conditions, where the proposed method effectively mitigates false alarms generated by the baseline network. In the second scenario, which involves nearshore small ships, it is evident that the baseline network experiences a higher rate of missed detections; conversely, our proposed method significantly reduces these missed detections. In the complex scenario of the third row, our proposed method not only effectively diminishes interference from rocks or structures—thereby reducing false alarms—but also prevents occurrences of missed detections altogether. The results indicate that our proposed method exhibits enhanced capabilities for detecting small ship targets while minimizing detection errors and improving overall detection accuracy.

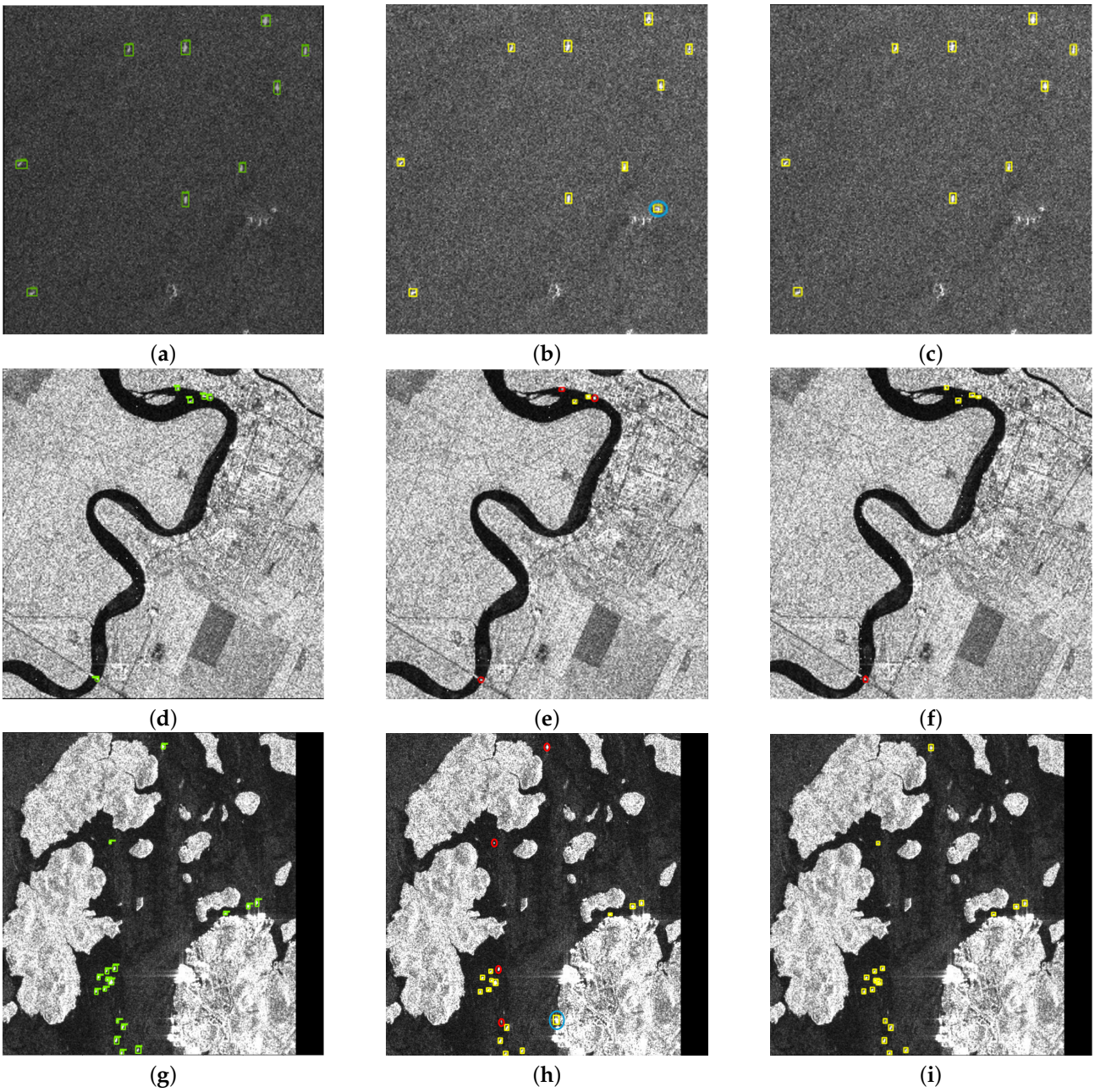


Figure 7. Visualization of the detection results. (**a,d,g**) represent the ground truth. (**b,e,h**) represent the baseline. (**c,f,i**) represent the proposed method. Green: ground truths. Yellow: detection results. Red: missed detections. Blue: false alarms.

Tables 3 and 4 show the results of the ablation experiments conducted on both the HRSID dataset and the iVision-MRSSD dataset. Given our emphasis on improving the detection rate of small targets, we have incorporated additional evaluation metrics—APs, APm, and APl—to assess the average precision of targets across different sizes, building upon the indicators established in Table 3. In terms of performance on the HRSID dataset, our proposed method demonstrates a notable improvement of 2.4% in both AP50 and AP, along with a 1.6% enhancement in APs. For the iVision-MRSSD dataset, our approach yields a 1.7% increase in AP and an improvement of 2.2% in APs. These experimental results highlight the effectiveness of the optimizations introduced in this study for enhancing SAR ship detection models.

Table 3. Results of ablation experiment on the HRSID dataset.

Method	SR	SPD	HA	Loss	P	R	AP50	AP	APs	APm	API
Baseline	—	—	—	—	90.2	82.7	88.6	64.3	64.7	74.5	13.4
	✓	—	—	—	90.5	82.6	89.7	65.3	65.0	73.9	14.7
	—	✓	—	—	90.2	82.6	89.7	66.0	65.2	74.8	23.0
	—	—	✓	—	90.4	82.7	89.6	65.7	64.8	74.9	19.7
	—	—	—	✓	89.6	83.6	89.2	64.5	64.7	74.9	17.8
	✓	✓	—	—	90.0	82.6	89.1	65.5	66.0	74.3	11.8
	✓	✓	✓	—	**90.9**	83.4	90.3	66.2	65.6	**76.0**	20.9
Proposed	✓	✓	✓	✓	90.4	**83.7**	**91.0**	**66.7**	**66.3**	74.5	**23.1**

Table 4. Results of ablation experiment on the iVision-MRSSD dataset.

Method	SR	SPD	HA	Loss	P	R	AP50	AP	APs	APm	API
Baseline	—	—	—	—	89.8	87.1	94.2	59.3	52.7	68.3	75.2
	✓	—	—	—	90.1	89.1	94.7	59.4	52.6	68.0	77.1
	—	✓	—	—	90.5	88.6	94.7	59.6	52.8	68.3	77.0
	—	—	✓	—	89.9	88.7	94.4	59.3	52.2	68.3	76.4
	—	—	—	✓	90.6	87.5	94.5	59.7	53.1	68.3	76.1
	✓	✓	—	—	91.2	89.1	95.0	60.1	53.7	68.1	76.9
	✓	✓	✓	—	**91.7**	88.9	95.1	60.2	53.8	68.7	77.1
Proposed	✓	✓	✓	✓	90.4	**89.8**	**95.2**	**61.0**	**54.9**	**68.8**	**78.3**

To demonstrate the universality of the proposed method, detection results from a variety of SAR satellite images are presented in Figure 8. The first row of images depicts small ship targets against a noisy background. It is evident that the baseline network experiences both missed and false detections, whereas the proposed method successfully mitigates these issues. Furthermore, the proposed method effectively identifies omissions made by the baseline network in scenes featuring small targets amidst man-made construction interference near shorelines, as shown in the second row. The third and fourth rows illustrate medium-sized target scenarios within similar near-shore environments, while the final row shows large target scenes. Notably, the proposed method rectifies problems associated with false alarms and inaccurate detection bounding box observed in the baseline network when subjected to interference. In general, scenarios involving small targets exhibit a higher likelihood of missed detections; conversely, for multi-scale ship targets in nearshore contexts, false alarms tend to be more prevalent due to disruptions caused by coastal infrastructure. Overall, our proposed method demonstrates superior performance compared to the baseline network across these diverse scenarios.

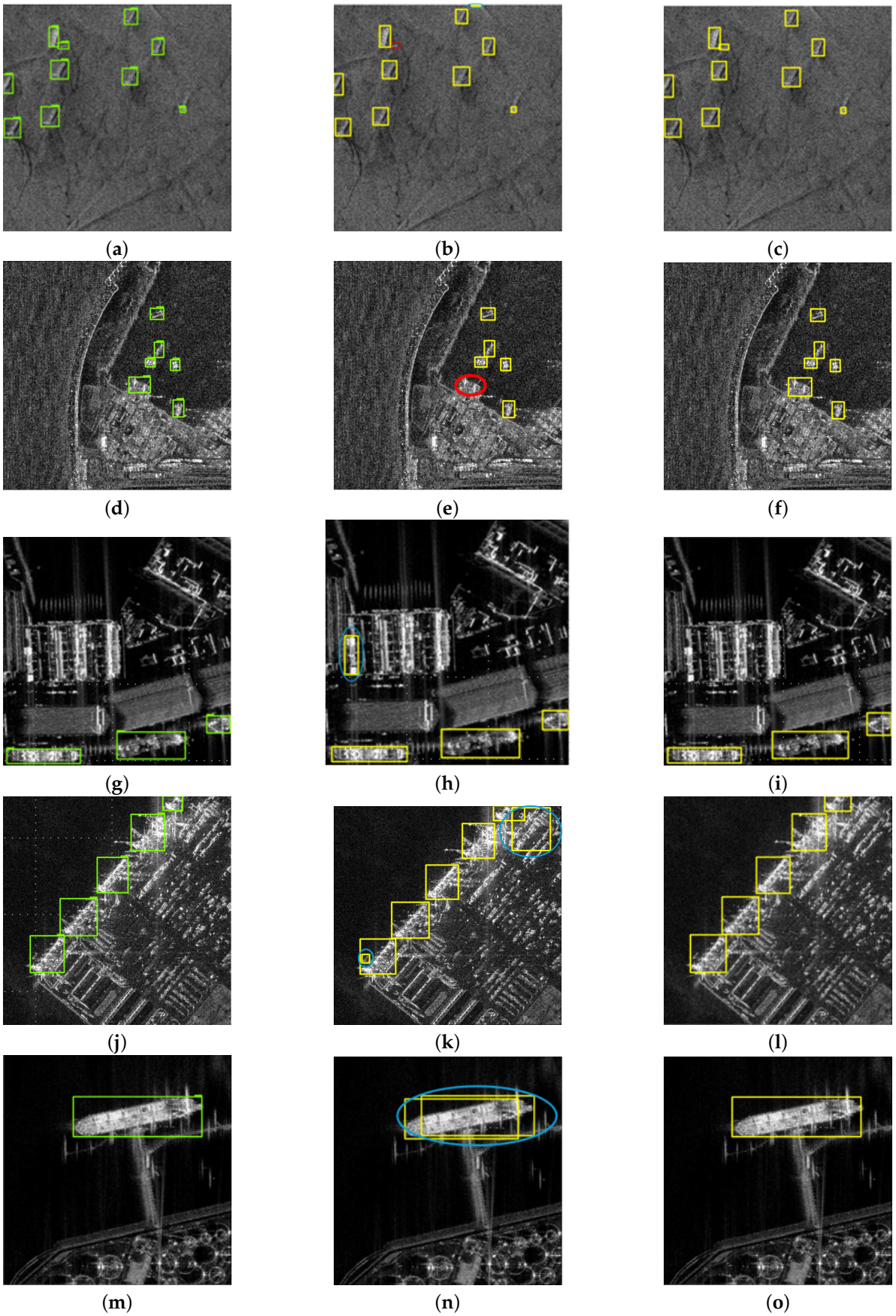


Figure 8. Visualization of the detection results. (**a,d,g,j,m**) represent the ground truth. (**b,e,h,k,n**) represent the baseline. (**c,f,i,l,o**) represent the proposed method. Green: ground truths. Yellow: detection results. Red: missed detections. Blue: false alarms.

3.4. Comparison Experiments

To further evaluate the effectiveness of the proposed method, this section compares it with several other commonly used target detection methods on the LS-SSDD dataset, HRSID dataset and iVision-MRSSD dataset. We employed eight general object detection algorithms, namely Faster RCNN [17], CenterNet [50], FCOS [51], ATSS [52], YOLOv5n, YOLOv8n, YOLOv10n [22], and YOLOv11n. Additionally, we included two algorithms specifically designed for ship detection in SAR images: SHIP-YOLO [28] and LHSDNet [29], as part of our comparative experiments. A quantitative analysis was conducted on the three datasets utilized in this study. The detailed detection results are presented in Table 5.

Table 5. Results of comparison experiments.

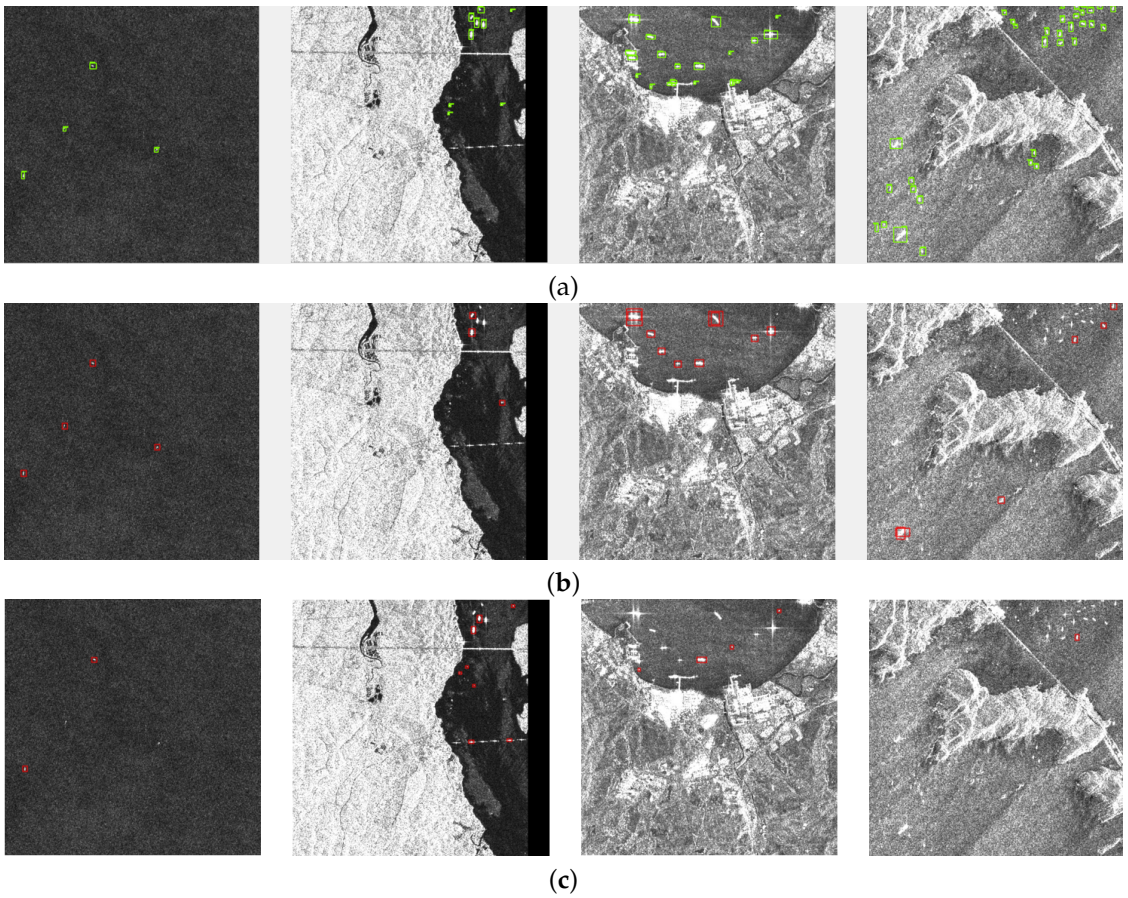
	LS-SSDD		**HRSID**					**iVision-MRSSD**				
Method	**AP50**	**AP**	**AP50**	**AP**	**APs**	**APm**	**API**	**AP50**	**AP**	**APs**	**APm**	**API**
Faster R-CNN	56.6	17.2	78.2	53.2	54.2	28.0	2.4	91.2	57.7	50.7	65.5	65.7
CenterNet	59.5	19.2	85.7	56.8	57.7	35.2	14.4	92.9	59.1	51.9	66.1	71.8
FCOS	59.6	19.6	69.9	41.5	43.1	7.6	2.8	78.9	43.5	39.3	50.1	39.6
ATSS	62.6	21.0	77.0	49.2	50.3	23.9	8.5	89.6	53.2	46.7	61.4	59.2
yolov5n	69.5	27.6	86.3	61.7	61.3	69.1	8.3	91.9	57.5	51.1	66.4	69.5
yolov8n	71.6	27.4	87.7	62.7	62.1	71.3	11.5	92.7	58.1	51.5	66.7	76.3
yolov10n	68.8	26.9	83.7	58.8	59.2	61.7	7.4	91.6	57.2	51.6	66.0	64.5
yolov11n	70.1	26.8	86.3	61.7	60.9	69.9	9.0	91.8	57.9	52.1	66.1	69.6
SHIP-YOLO	70.8	27.1	86.0	61.3	62.5	70.2	7.7	93.1	58.7	52.4	66.8	71.7
LHSDNet	72.2	27.8	87.0	60.7	60.9	69.1	12.1	93.7	58.3	52.0	66.8	74.8
Proposed	**73.7**	**29.2**	**88.4**	**63.4**	**62.9**	**72.4**	**15.4**	**94.1**	**59.6**	**53.1**	**67.9**	**76.4**

Notably, the average precision values for the latest YOLOv10n and YOLOv11n in SAR image ship detection were found to be lower than those of our baseline network, YOLOv8n. This discrepancy may stem from the influence of data characteristics and application scenarios on model performance. While both YOLOv10 and YOLOv11 demonstrate superior performance in general object detection tasks, it appears that YOLOv8 exhibits greater flexibility in complex scenes and achieves better results specifically for SAR image ship detection. Furthermore, it is crucial to highlight that the two networks tailored for SAR image ship detection demonstrate good performance on only one or two datasets but fail to achieve satisfactory results across all three datasets simultaneously. In contrast, the method proposed in this paper exhibits outstanding performance across all three datasets.

The HRSID dataset accounts for over 90% of small targets, while the LS-SSDD dataset comprises nearly all small targets. It is evident that only the YOLO series algorithms perform effectively on these two datasets, whereas other algorithms exhibit low AP values. This observation further underscores the advantages of YOLO series algorithms in detecting small targets. The proposed method optimizes upon YOLOv8n, achieving a 2% improvement on the LS-SSDD dataset and a 1% enhancement on the HRSID dataset. The APl is low and irregular in HRSID due to the presence of less than 1% of large targets within the dataset. In terms of the comprehensive iVision-MRSSD dataset, both Faster R-CNN and CenterNet demonstrate strong performance. However, when it comes to small target detection, our proposed method achieves an additional 2% increase in AP value. Whether evaluated by average precision at an IoU threshold of 0.5 (AP50) or by small target measurement index APs, our method consistently attains the highest values. This outcome underscores the effectiveness and feasibility of our approach.

Our comparative experiments encompass eight general-purpose target detection networks and two SAR ship detection networks. For qualitative analysis, we selected various scenes from the LS-SSDD dataset, including maritime environments, nearshore settings, and complex background scenarios. The results are illustrated in Figure 9. The first column presents sparse small ship targets located in the far sea. Both Faster R-CNN and YOLO se-

ries methods, along with SAR image ship detection algorithms, demonstrate commendable performance in this context. The second column depicts a nearshore scene featuring small targets. Conventional target detection algorithms exhibit significant missed detections. Although CenterNet successfully identifies most small targets, it concurrently generates numerous false alarms. In contrast, LHSDNet and the proposed method show superior performance. In terms of complex scenes represented in the third column—where multiple ships coexist alongside islands, reefs, offshore structures, and other disturbances—various methods display varying degrees of missed detections and false alarms. Notably, our proposed method outperforms others by detecting a greater number of targets. This underscores its effectiveness in identifying sparse small targets within intricate environments. However, as depicted in the fourth column where background noise is further intensified, all methods experience substantial degradation in performance. Even specialized algorithms for SAR image ship detection struggle to identify ship targets under these conditions; nevertheless, our proposed method continues to outperform alternative approaches by successfully detecting over half of the visible ship targets presented in the figure. This series of detection outcomes validate the efficacy of our improved methodology and highlights its robust capability for detecting small ship targets across diverse scenarios.



(a)

(**b**)

(**c**)

Figure 9. *Cont.*

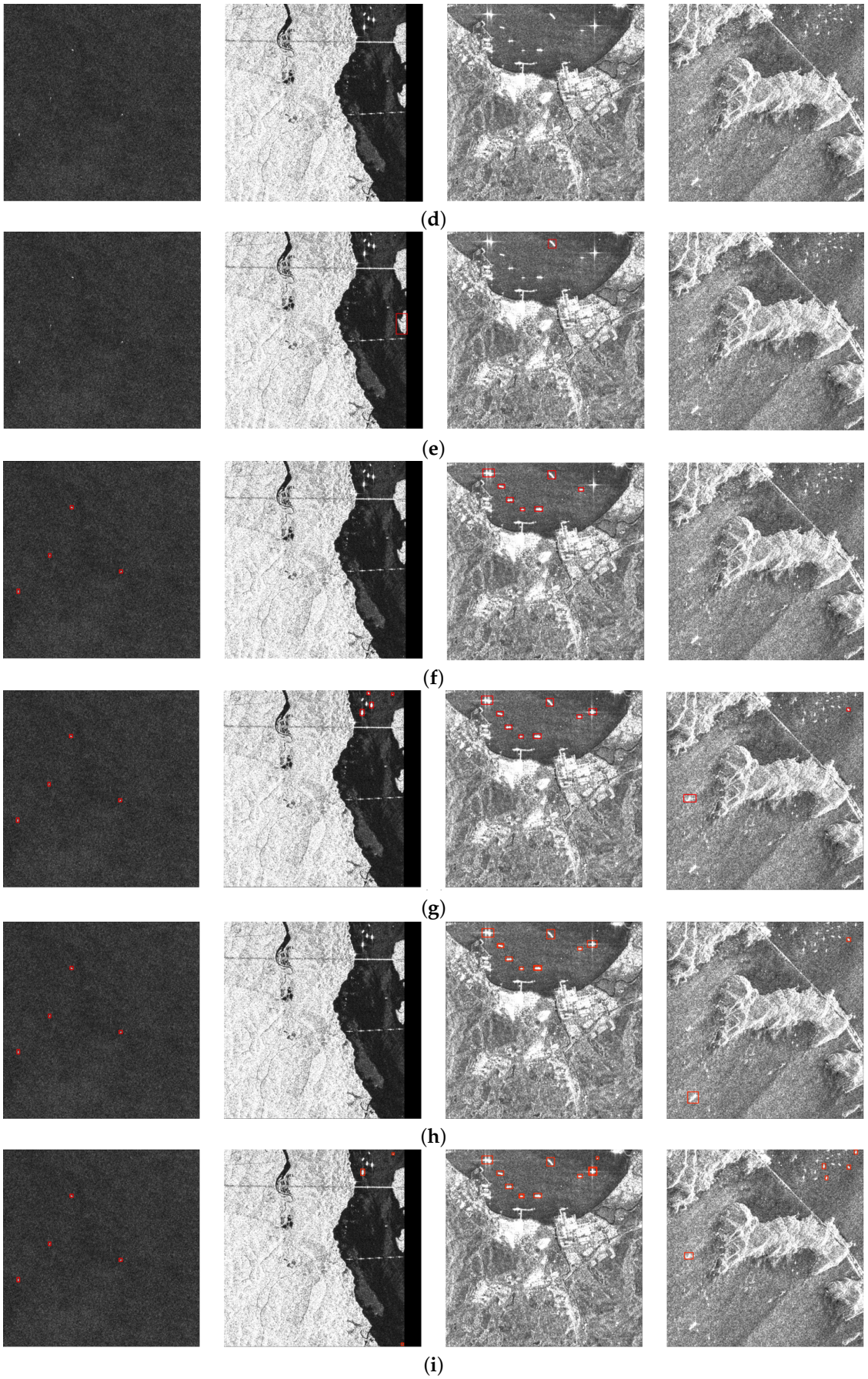


Figure 9. *Cont.*

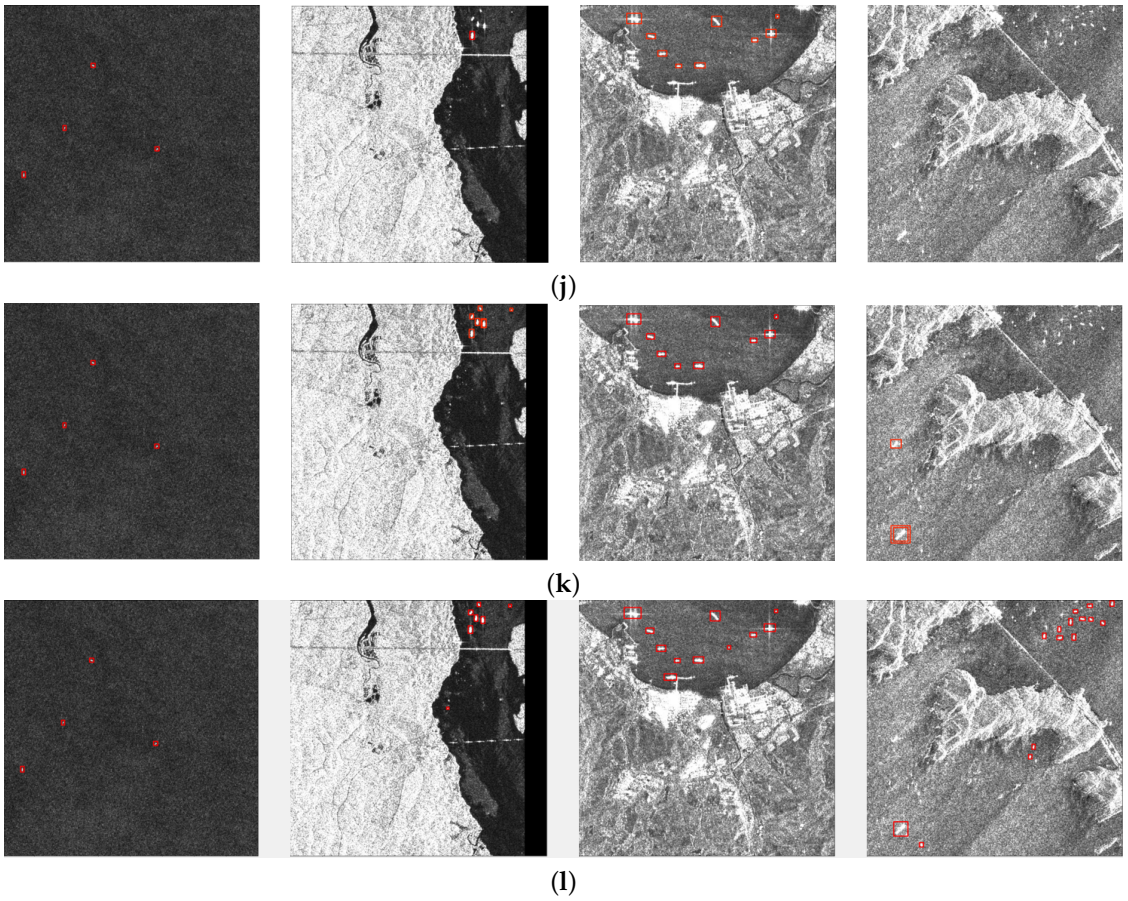


Figure 9. Visualization of the detection results of various methods on LS-SSDD (**a**)–(**l**). (**a**) Ground truth. (**b**) Faster R-CNN. (**c**) CenterNet. (**d**) FCOS. (**e**) ATSS. (**f**) YOLOv5n. (**g**) YOLOv8n. (**h**) YOLOv10n. (**i**) YOLOv11n. (**j**) SHIP-YOLO. (**k**) LHSDNet. (**l**) The proposed method. The green boxes indicate the ground truths; the red boxes indicate detection results.

4. Discussion

Extensive experimental results based on three public datasets have demonstrated the superior performance of the proposed method for small-sized ship detection when compared to other object detection techniques. The LS-SSDD dataset is specifically designed for detecting small-sized ships within large scenes, where small targets account for over 99.8% of the total. The proposed method achieves the highest Average Precision at IoU threshold 50 (AP50) and overall Average Precision (AP) values, surpassing the baseline YOLOv8n by 0.021 and 0.018, respectively. For the HRSID and iVision-MRSSD datasets, we focus on APs metrics other than AP50 and AP, which characterize the average precision of small targets. In comparison to the baseline YOLOv8n, our proposed method enhances these APs metrics by 0.022 on the iVision-MRSSD dataset. Given the diversity present in the iVision-MRSSD dataset, we can conclude that our proposed method is effective in improving detection performance for small-sized ships.

The enhancement of the proposed method primarily stems from improvements made to the backbone, neck, and loss function, tailored to the characteristics of ships within the datasets. The backbone incorporates both SR and SPD modules, which effectively mitigate the issue of feature disappearance for small vessels while ensuring comprehensive feature extraction. An attention module (HA) has been integrated into the neck, enabling effective suppression of background noise and facilitating feature fusion across different levels. Furthermore, the loss function is designed to account for the sensitivity of small ships to variations in Intersection over Union (IoU), as well as incorporating shape and

scale factors related to bounding boxes. This approach significantly enhances detection accuracy. Experimental results indicate that enhancements for small ships are considerably more pronounced compared to those for larger vessels; given that small ships constitute a substantial proportion of the datasets, these strategies collectively contribute to an overall improvement in performance.

In addition to detection accuracy, the computational efficiency of the detection network constitutes a critical factor. Table 6 presents a comparison of the computational efficiency and inference time between the baseline network YOLOv8n and the proposed method. To enhance the detection accuracy for small ship targets, we have increased both the model complexity and the resource requirements for training the network. Nevertheless, owing to the implementation of the re-parameterization module, the time consumption during the inference stage remains within a manageable range.

Table 6. Comparison of computational efficiency and inference speed.

Method	Params/M	FLOPs/G	Speed/ms
Baseline	3.0	8.1	3.5
Proposed	8.5	19.2	5.4

When inferring real SAR images, the initial step involves segmenting the large image into sub-images that match the sizes of the training data. Subsequently, ship detection is performed on these sub-images. Following detection, the sub-images are reassembled to reconstruct the complete SAR image as required. This approach is necessitated by the fact that in SAR image ship detection, ship size is determined based on the proportion of pixels occupied by each target. To ensure optimal performance of the detection network, it is essential that test images entering the network maintain pixel sizes consistent with those of the training images. As an illustrative example, we consider reasoning applied to the 11th image from the LS-SSDD dataset. The original image measuring 24,000 $\times$ 16,000 pixels was divided into 600 sub-images of 800 $\times$ 800 pixels each. The inference time for processing each sub-image was recorded at 5.4 ms, while the reasoning time for the whole image was 5 s, including data loading, image segmentation, and subsequent splicing operations.

For future work, we have the following considerations. Firstly, the datasets utilized in this study consist solely of horizontal bounding boxes; thus, there is potential to enhance the detection head for application in ship detection datasets derived from SAR images that utilize rotated bounding boxes. Secondly, given the real-time requirements associated with ship detection in SAR imagery, it is imperative to take into account both network complexity and computational resource consumption. In our future endeavors, we will explore lightweight modifications to the network architecture while ensuring that detection accuracy is preserved. Finally, we aim to investigate the integration of the proposed module into other network structures to enhance detection accuracy across a broader range of target detection tasks.

5. Conclusions

In this article, we propose an enhanced YOLO network featuring a redesigned backbone, neck, and loss function specifically tailored for the detection of small-sized ships. The SR module is employed to facilitate advanced and complete feature extraction and feature fusion while effectively managing parameter costs. The SPD module performs down-sampling operations and minimizes information loss. Additionally, the HA module integrates spatial and frequency domain information to enhance effective features while mitigating sea clutter and noise interference. Furthermore, the shape-NWD loss emphasizes the shape and scale of bounding boxes, thereby reducing the sensitivity of IoU to

positional deviations in small-sized ship detections. The ablation experiments illustrate the effectiveness of these components. In comparison experiments, our method demonstrates superior performance compared to other CNN approaches, including the state-of-the-art target detection algorithm YOLOv11 and algorithms specially designed for SAR image ship detection. The extensive experimental results provide compelling evidence that the proposed method demonstrates robust detection capabilities in identifying small ships within SAR images.

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

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–43. [CrossRef]
2. Chang, S.; Deng, Y.; Zhang, Y.; Zhao, Q.; Wang, R.; Zhang, K. An advanced scheme for range ambiguity suppression of spaceborne SAR based on blind source separation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5230112. [CrossRef]
3. Chang, S.; Deng, Y.; Zhang, Y.; Wang, R.; Qiu, J.; Wang, W.; Zhao, Q.; Liu, D. An advanced echo separation scheme for space-time waveform-encoding SAR based on digital beamforming and blind source separation. *Remote Sens.* **2022**, *14*, 3585. [CrossRef]
4. Liu, D.; Chang, S.; Deng, Y.; He, Z.; Wang, F.; Zhang, Z.; Han, C.; Yu, C. A Novel Spaceborne SAR Constellation Scheduling Algorithm for Sea Surface Moving Target Search Tasks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 3715–3726. [CrossRef]
5. Bürgmann, R.; Rosen, P.A.; Fielding, E.J. Synthetic aperture radar interferometry to measure Earth's surface topography and its deformation. *Annu. Rev. Earth Planet. Sci.* **2000**, *28*, 169–209. [CrossRef]
6. Liu, C.A.; Chen, Z.X.; Yun, S.; Chen, J.S.; Hasi, T.; Pan, H.Z. Research advances of SAR remote sensing for agriculture applications: A review. *J. Integr. Agric.* **2019**, *18*, 506–525. [CrossRef]
7. Sun, Z.; Leng, X.; Zhang, X.; Xiong, B.; Ji, K.; Kuang, G. Ship recognition for complex SAR images via dual-branch transformer fusion network. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 4009905. [CrossRef]
8. Gens, R. Oceanographic applications of SAR remote sensing. *Gisci. Remote Sens.* **2008**, *45*, 275–305. [CrossRef]
9. Gao, G. Statistical modeling of SAR images: A survey. *Sensors* **2010**, *10*, 775–795. [CrossRef]
10. Renga, A.; Graziano, M.D.; Moccia, A. Segmentation of marine SAR images by sublook analysis and application to sea traffic monitoring. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 1463–1477. [CrossRef]
11. Li, L.; Du, L.; Wang, Z. Target detection based on dual-domain sparse reconstruction saliency in SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4230–4243. [CrossRef]
12. Schwegmann, C.P.; Kleynhans, W.; Salmon, B.P. Synthetic aperture radar ship detection using Haar-like features. *IEEE Geosci. Remote Sens. Lett.* **2016**, *14*, 154–158. [CrossRef]
13. Tao, L.; Ziyuan, Y.; Yanni, J.; Gui, G. Review of ship detection in polarimetric synthetic aperture imagery. *J. Radar* **2021**, *10*, 1–19.
14. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
15. Li, J.; Xu, C.; Su, H.; Gao, L.; Wang, T. Deep learning for SAR ship detection: Past, present and future. *Remote Sens.* **2022**, *14*, 2712. [CrossRef]
16. Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object detection in 20 years: A survey. *Proc. IEEE* **2023**, *111*, 257–276. [CrossRef]
17. 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.* **2016**, *39*, 1137–1149. [CrossRef]
18. Redmon, J. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 25 June–1 July 2016.

19. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
20. Ge, Z. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
21. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
22. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. Yolov10: Real-time end-to-end object detection. *Adv. Neural Inf. Process. Syst.* **2025**. *37*, 107984–108011.
23. Cui, Z.; Wang, X.; Liu, N.; Cao, Z.; Yang, J. Ship detection in large-scale SAR images via spatial shuffle-group enhance attention. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 379–391. [CrossRef]
24. Shan, H.; Fu, X.; Lv, Z.; Zhang, Y. SAR ship detection algorithm based on deep dense sim attention mechanism network. *IEEE Sens. J.* **2023**, *23*, 16032–16041. [CrossRef]
25. Hu, Q.; Hu, S.; Liu, S. BANet: A balance attention network for anchor-free ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5222212. [CrossRef]
26. Sun, Z.; Leng, X.; Lei, Y.; Xiong, B.; Ji, K.; Kuang, G. BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images. *Remote Sens.* **2021**, *13*, 4209. [CrossRef]
27. Liangjun, Z.; Feng, N.; Yubin, X.; Gang, L.; Zhongliang, H.; Yuanyang, Z. MSFA-YOLO: A Multi-Scale SAR Ship Detection Algorithm Based on Fused Attention. *IEEE Access* **2024**, *12*, 24554–24568. [CrossRef]
28. Luo, Y.; Li, M.; Wen, G.; Tan, Y.; Shi, C. SHIP-YOLO: A Lightweight Synthetic Aperture Radar Ship Detection Model based on YOLOv8n Algorithm. *IEEE Access* **2024** , *12*, 37030–37041. [CrossRef]
29. Dai, D.; Wu, H.; Wang, Y.; Ji, P. LHSDNet: A Lightweight and High-Accuracy SAR Ship Object Detection Algorithm. *Remote Sens.* **2024**, *16*, 4527. [CrossRef]
30. Lanz, P.; Marino, A.; Brinkhoff, T.; Köster, F.; Möller, M. The InflateSAR campaign: Testing SAR vessel detection systems for refugee rubber inflatables. *Remote Sens.* **2021**, *13*, 1487. [CrossRef]
31. Lanz, P.; Marino, A.; Simpson, M.D.; Brinkhoff, T.; Köster, F.; Möller, M. The InflateSAR campaign: Developing refugee vessel detection capabilities with polarimetric SAR. *Remote Sens.* **2023**, *15*, 2008. [CrossRef]
32. Zakharov, I.; Henschel, M.D.; Power, D.; Burke, P.; Puestow, T.; Warren, S. Signatures of Small Boats With TerraSAR-X Staring Spotlight Data. *IEEE Geosci. Remote Sens. Lett.* **2022**, *20*, 4001005. [CrossRef]
33. Shin, D.W.; Yang, C.S.; Chowdhury, S.J.K. Enhancement of Small Ship Detection Using Polarimetric Combination from Sentinel- 1 Imagery. *Remote Sens.* **2024**, *16*, 1198. [CrossRef]
34. Wang, J.; Lin, Y.; Guo, J.; Zhuang, L. SSS-YOLO: Towards more accurate detection for small ships in SAR image. *Remote Sens. Lett.* **2021**, *12*, 93–102. [CrossRef]
35. Yang, S.; An, W.; Li, S.; Wei, G.; Zou, B. An improved FCOS method for ship detection in SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8910–8927. [CrossRef]
36. Zhou, H.; Chen, P.; Li, Y.; Wang, B. Enhanced Detection Method for Small and Occluded Targets in Large-Scene Synthetic Aperture Radar Images. *J. Mar. Sci. Eng.* **2023**, *11*, 2081. [CrossRef]
37. Hu, B.; Miao, H. An Improved Deep Netural Network for Small Ship Detection in SAR Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *17*, 2594–2609.
38. Ding, X.; Zhang, X.; Han, J.; Ding, G. Diverse branch block: Building a convolution as an inception-like unit. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10886–10895.
39. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.
40. Lee, Y.; Park, J. Centermask: Real-time anchor-free instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 13906–13915.
41. Sunkara, R.; Luo, T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Grenoble, France, 19–23 September 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 443–459.
42. Ruan, J.; Gao, J.; Xie, M.; Xiang, S. Learning Multi-axis Representation in Frequency Domain for Medical Image Segmentation. *arXiv* **2023**, arXiv:2312.17030. [CrossRef]
43. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
44. Zhang, H.; Zhang, S. Shape-iou: More accurate metric considering bounding box shape and scale. *arXiv* **2023**, arXiv:2312.17663.
45. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part V 13; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.

46. Zhang, T.; Zhang, X.; Ke, X.; Zhan, X.; Shi, J.; Wei, S.; Pan, D.; Li, J.; Su, H.; Zhou, Y.; et al. LS-SSDD-v1. 0: A deep learning dataset dedicated to small ship detection from large-scale Sentinel-1 SAR images. *Remote Sens.* **2020**, *12*, 2997. [CrossRef]
47. Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; Shi, J. HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation. *IEEE Access* **2020**, *8*, 120234–120254. [CrossRef]
48. Humayun, M.F.; Bhatti, F.A.; Khurshid, K. iVision MRSSD: A comprehensive multi-resolution SAR ship detection dataset for state of the art satellite based maritime surveillance applications. *Data Brief* **2023**, *50*, 109505. [CrossRef]
49. Chen, K.; Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Xu, J.; et al. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv* **2019**, arXiv:1906.07155.
50. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 Ocotber–2 November 2019; pp. 6569–6578.
51. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: A simple and strong anchor-free object detector. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 1922–1933. [CrossRef] [PubMed]
52. Zhang, S.; Chi, C.; Yao, Y.; Lei, Z.; Li, S.Z. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 9759–9768.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.