remote sensing

MDPI

Article

HLNet: A Lightweight Network for Ship Detection in Complex SAR Environments

Xiaopeng Guo [1], Fan Deng [1,*], Jie Gong [2], Jing Zhang [3], Jiajia Guo [1], Yong Wang [1], Yinmei Zeng [1] and Gongquan Li [1]

[1] School of Geosciences, Yangtze University, Wuhan 430100, China
[2] Wuhan Huaxin Lianchuang Technology Engineering Co., Ltd., Wuhan 430074, China
[3] The 9th Geological Brigade, Hebei Bureau of Geology and Mineral Resources Exploration, Xingtai 054000, China
* Correspondence: dengfan@yangtzeu.edu.cn

Highlights

What are the main findings?

- A high-precision lightweight SAR ship detection network (HLNet) is proposed, achieving mAP_{50} of 98.3% on SSDD and 91.7% on HRSID with only 0.66 M parameters.
- The designed MSNet–HLPAFPN–HLHead architecture effectively improves multi-scale ship perception and localization robustness under strong speckle noise and complex sea clutter.

What are the implications of the main findings?

- The proposed lightweight design establishes a new accuracy–efficiency trade-off for SAR ship detection, outperforming existing detectors under extreme parameter constraints.
- HLNet is well-suited for deployment on resource-constrained SAR platforms, enabling reliable ship detection for satellite-borne, airborne, and embedded edge systems.

Abstract

The coherent speckle noise in synthetic aperture radar (SAR) imagery, together with complex sea clutter and large variations in ship target scales, poses significant challenges to accurate and robust ship detection, particularly under strict lightweight constraints required by satellite-borne and airborne platforms. To address this issue, this paper proposes a high-precision lightweight detection network, termed High-Lightweight Net (HLNet), specifically designed for SAR ship detection. The network incorporates a novel multi-scale backbone, Multi-Scale Net (MSNet), which integrates dynamic feature completion and multi-core parallel convolutions to alleviate small-target feature loss and suppress background interference. To further enhance multi-scale feature fusion while reducing model complexity, a lightweight path aggregation feature pyramid network, High-Lightweight Feature Pyramid (HLPAFPN), is introduced by reconstructing fusion pathways and removing redundant channels. In addition, a lightweight detection head, High-Lightweight Head (HLHead), is designed by combining grouped convolutions with distribution focal loss to improve localization robustness under low signal-to-noise ratio conditions. Extensive experiments conducted on the public SSDD and HRSID datasets demonstrate that HLNet achieves mAP_{50} scores of 98.3% and 91.7%, respectively, with only 0.66 M parameters. Extensive evaluations on the more challenging CSID subset, composed of complex scenes selected from SSDD and HRSID, demonstrate that HLNet attains an mAP_{50} of 75.9%, outperforming the baseline by 4.3%. These results indicate that HLNet achieves an effective

Check for updates

Academic Editor: Kefeng Ji

Received: 23 January 2026
Revised: 7 February 2026
Accepted: 10 February 2026
Published: 12 February 2026

Copyright: © 2026 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.

balance between detection accuracy and computational efficiency, making it well-suited for deployment on resource-constrained SAR platforms.

Keywords: ship detection; synthetic aperture radar; lightweight network; multi-scale feature fusion; edge deployment

1. Introduction

Maritime ship detection holds irreplaceable strategic value in fields such as ocean traffic monitoring [1], fisheries management [2,3], and military reconnaissance [4]. Synthetic Aperture Radar (SAR), with its all-weather, all-time operational capability [5], is widely considered the preferred method for wide-area ocean surveillance [6]. However, practical application scenarios present two significant challenges to SAR-based intelligent interpretation algorithms. On the one hand, the coherent speckle noise in SAR imaging mechanisms [7,8], coupled with complex and variable coastal backgrounds and the wide size range of ship targets [9], demands models with exceptional feature extraction and anti-interference capabilities. On the other hand, the extremely limited power and computational resources of edge computing platforms, such as satellite-borne and airborne systems, urgently require detection models to undergo extreme lightweight optimization [10,11].

Traditional SAR ship detection methods primarily rely on constant false alarm rate (CFAR) [12] and its variant algorithms, such as CA-CFAR [13], GO-CFAR [14], and VI-CFAR [15]. These approaches rely on manually designed statistical distribution models to characterize sea clutter [16]. While effective under simple and uniform sea conditions, these methods often exhibit limited generalization capability and significantly elevated false alarm rates when confronted with non-uniform clutter (e.g., busy coastal areas), due to difficulties in accurately fitting complex background statistical characteristics [17].

In recent years, deep learning techniques have achieved significant breakthroughs in SAR ship detection due to their powerful data-driven feature representation capabilities [18], including generic detection networks and task-driven models that exploit scattering characteristics or domain-specific priors [19,20]. Early studies often directly adapted mature frameworks from natural image detection domains, such as the two-stage Faster R-CNN [21] and single-stage approaches like the Single Shot MultiBox Detector (SSD) [22] and You Only Look Once (YOLO) [23] series. Two-stage methods offer higher accuracy but suffer from substantial computational redundancy, making it difficult for them to meet real-time requirements. Single-stage methods, being faster and lighter, have gradually become the dominant choice for current tasks [24,25].

However, directly applying generic lightweight single-stage detectors to SAR ship detection remains highly challenging, especially under complex maritime scenarios. First, to pursue computational efficiency, most lightweight detectors adopt aggressive downsampling strategies. While effective for optical imagery, this design is problematic for SAR ship detection, where small vessels often occupy only a few pixels in high-resolution SAR images. As a result, fine-grained spatial information of small ships is easily lost in deep layers, leading to missed detections. Second, existing lightweight backbones are primarily designed with efficiency-oriented convolutional operations (e.g., depthwise or grouped convolutions) and lack mechanisms explicitly tailored to suppress coherent speckle noise inherent in SAR imaging. In complex backgrounds such as near-shore or port areas, this limitation significantly degrades feature discriminability and increases false alarms. Third, commonly used feature pyramid structures, such as FPN and PANet, were originally developed for generic object detection tasks with diverse and balanced scale distributions.

However, SAR ship targets exhibit a highly skewed scale distribution, with dense medium- and small-scale targets dominating most scenes. Consequently, large-scale pyramid layers contribute marginally while introducing substantial channel redundancy and unnecessary computational overhead, which is particularly unfavorable for resource-constrained edge platforms.

In summary, existing lightweight detectors, although efficient on natural images, often fail to achieve a favorable accuracy–efficiency trade-off for SAR ship detection in complex maritime environments [26]. Designing a SAR-aware lightweight architecture that simultaneously preserves small-target details, suppresses speckle noise, and eliminates redundant feature representations remains an open challenge.

To address the aforementioned challenges, this paper proposes a high-performance lightweight SAR ship detection network, termed HLNet. A key consideration in HLNet's design is the trade-off between lightness and detection precision. Each module is carefully designed to reduce computational overhead while preserving or enhancing critical features: the Multi-Scale Convolution (MSConv) and Feature Complementary Mapping (FCM) [27] modules retain fine-grained small-ship details and suppress speckle noise; the High-Lightweight Feature Pyramid (HLPAFPN) removes redundant channels and layers while maintaining effective multi-scale feature fusion; and the High-Lightweight Head (HLHead), together with Distribution Focal Loss (DFL) [28] and grouped convolutions, enhances boundary modeling and localization under fuzzy conditions. Through this coordinated design, HLNet achieves high accuracy on SAR ship detection tasks while operating efficiently on resource-constrained edge platforms, demonstrating that the coordinated lightweight strategies effectively balance computational efficiency and detection precision. The main contributions of this work are summarized as follows:

(1) We propose an efficient multi-scale backbone network, termed MSNet, to address the challenges of coherent speckle noise and multi-scale ship representation under strict computational constraints. Specifically, the MSConv module employs small kernels to suppress high-frequency speckle noise while using large kernels to enlarge the receptive field for medium-to-large ships. In addition, the FCM module is introduced to preserve the fine-grained spatial details of small ships that are easily lost in deep lightweight networks. Unlike the original FCM designed for small targets in aerial imagery, the FCM module, when used together with MSConv, further alleviates the degradation of small-ship features caused by strong speckle noise in SAR images, while maintaining computational efficiency.

(2) We design a SAR-customized lightweight feature pyramid, named HLPAFPN, to overcome the inefficiency of generic FPN/PAN structures in SAR ship detection [29]. By analyzing the scale distribution characteristics of SAR ship targets, HLPAFPN reconstructs the feature fusion path, removes redundant large-scale layers unsuitable for SAR scenes, and compresses non-critical channels. This design significantly reduces computational redundancy while maintaining effective multi-scale feature fusion.

(3) To improve localization robustness under fuzzy ship boundaries in SAR imagery, we propose a high-precision lightweight detection head, termed HLHead, which incorporates the DFL. Originally proposed for high-performance detectors, DFL's integration into the lightweight architecture enhances boundary modeling and localization accuracy, ensuring high precision despite reduced computational resources. Furthermore, we reveal a structural synergy between grouped convolutions and DFL by aligning the number of convolution groups with the number of discrete DFL intervals, resulting in a unified and computation-efficient design.

Extensive experiments demonstrate that HLNet achieves leading detection accuracy with minimal resource consumption. Its effectiveness is confirmed on both the Sentinel Ship

Detection Dataset (SSDD) [30] and the High-Resolution SAR Images Dataset (HRSID) [31]. This provides a practical and efficient solution for real-time intelligent SAR processing on resource-constrained platforms.

This paper is organized into six sections. Section 2 provides a comprehensive review of existing SAR ship detection methods based on deep learning approaches. Section 3 presents an analysis of the overall architecture of HLNet. Experimental results are presented in Section 4. Section 5 discusses the limitations of the proposed model and outlines future research directions. Finally, Section 6 concludes the paper.

2. Related Works

Deep learning-based SAR ship detection methods can be broadly categorized into two main types based on their workflow architecture: two-stage and one-stage methods. This section reviews relevant research from both perspectives.

Faster R-CNN and its variants (such as Mask R-CNN [32]), as representative high-precision two-stage models, have been widely adopted by researchers in the field of SAR ship target extraction [33,34]. Liu et al. proposed the two-stage anchor-free ship detection algorithm FEAD, which builds upon FCOS [35] to construct a fully convolutional anchor-free detection network (FEN). By utilizing the RPN from Faster R-CNN as the stage-one network, the proposed method enhances the model's representational capability for detection targets and reduces the search space, thereby improving detection accuracy. Zhou et al. [36] specifically designed a module integrating attention mechanisms with multi-scale feature fusion to address the common challenges of small targets and complex backgrounds in SAR images. This approach enhances the capture of critical information while suppressing noise interference, thereby improving detection performance. To fully leverage background information and enhance detection accuracy for multi-scale ships, Xia et al. proposed CRTransSar [37], which is based on the two-stage detection model Mask R-CNN. CRTransSar innovatively integrates the global context awareness of Transformers with the local feature representation capabilities of CNNs. This integration enhances the model's detection performance but also increases its complexity.

Due to the advantages of one-stage object detection models—such as high speed, simple architecture, and easier deployment—many researchers have chosen to conduct studies based on one-stage models in recent years [38–40]. To enhance detection capabilities for multi-scale ships, Wei et al. leveraged the YOLOv3 framework. By incorporating a reference residual module structure to prevent network degradation and integrating feature maps of different scales from both deep and early layers, the proposed method achieved improved anti-interference performance and detection accuracy for multi-scale SAR ship detection. To further simplify the model and reduce reliance on predefined anchor boxes, Li et al. [41] proposed the TKP-Net three-keypoint detection network within an anchor-box-free framework. By detecting key points such as the bow and stern, the proposed network not only achieves ship localization but also determines its heading and course. Zhang et al. proposed EsarDet [42], which further enhances detection accuracy and computational efficiency. Based on YOLOX-Tiny, EsarDet incorporates the CAA-Net module to expand the field of view, thereby integrating contextual and semantic information from ships. Additionally, EsarDet introduces a novel spatial pyramid pooling network, A2SPPF, and proposes a new convolutional module, A2CSPlayer, to minimize the loss of target semantic information and enhance the fusion of diverse features. To meet the real-time processing demands of spaceborne and airborne platforms, model lightweighting has emerged as a research hotspot. For example, the BESW-YOLO [43] model demonstrates that careful architectural design can significantly reduce parameters while maintaining

high detection accuracy on SAR ship datasets, making it well-suited for deployment in resource-constrained scenarios.

In summary, SAR ship detection technology is evolving toward greater intelligence, precision, and efficiency. Building upon the work of previous researchers, this study aims to simultaneously enhance detection accuracy, computational efficiency, and real-time performance under stringent lightweight constraints.

3. Methods

3.1. Overall Network Structure

HLNet aims to achieve the dual objectives of high accuracy and extremely lightweight design. To this end, we adopted the YOLO11 as our foundation model, leveraging its efficiency as a basis for deep customization and pruning. As the recent iteration of the YOLO series, YOLO11 delivers significant improvements in model architecture and efficiency. Its core innovation lies in replacing the traditional C2f module with the C3k2 module, while introducing the C2PSA module as an efficient spatial attention mechanism inserted after the SPPF module.

The C3k2 module enhances feature reuse capability while reducing computational load through cross-layer connections and efficient convolutional design. The SPPF module significantly expands the model's receptive field via serial pooling and fusion, enabling more effective capture of multi-scale contextual information. Figure 1 illustrates the workflow structures of the SPPF and C3k2 modules.

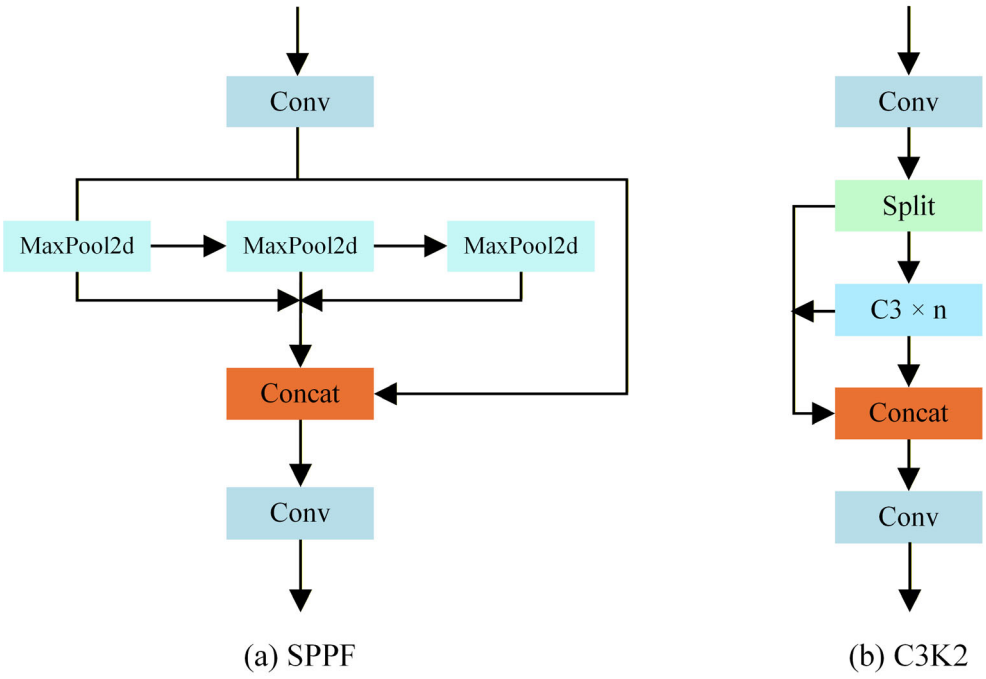


Figure 1. The specific structure diagrams. (**a**) The C3k2 structure diagrams. (**b**) The SPPF structure diagrams.

As illustrated in Figure 2, to achieve maximal efficiency, HLNet first eliminated the computationally intensive C2PSA module from the original YOLO11. Building upon this foundation, we introduced three core components: the MSNet, the HLPAFPN, and the HLHead.

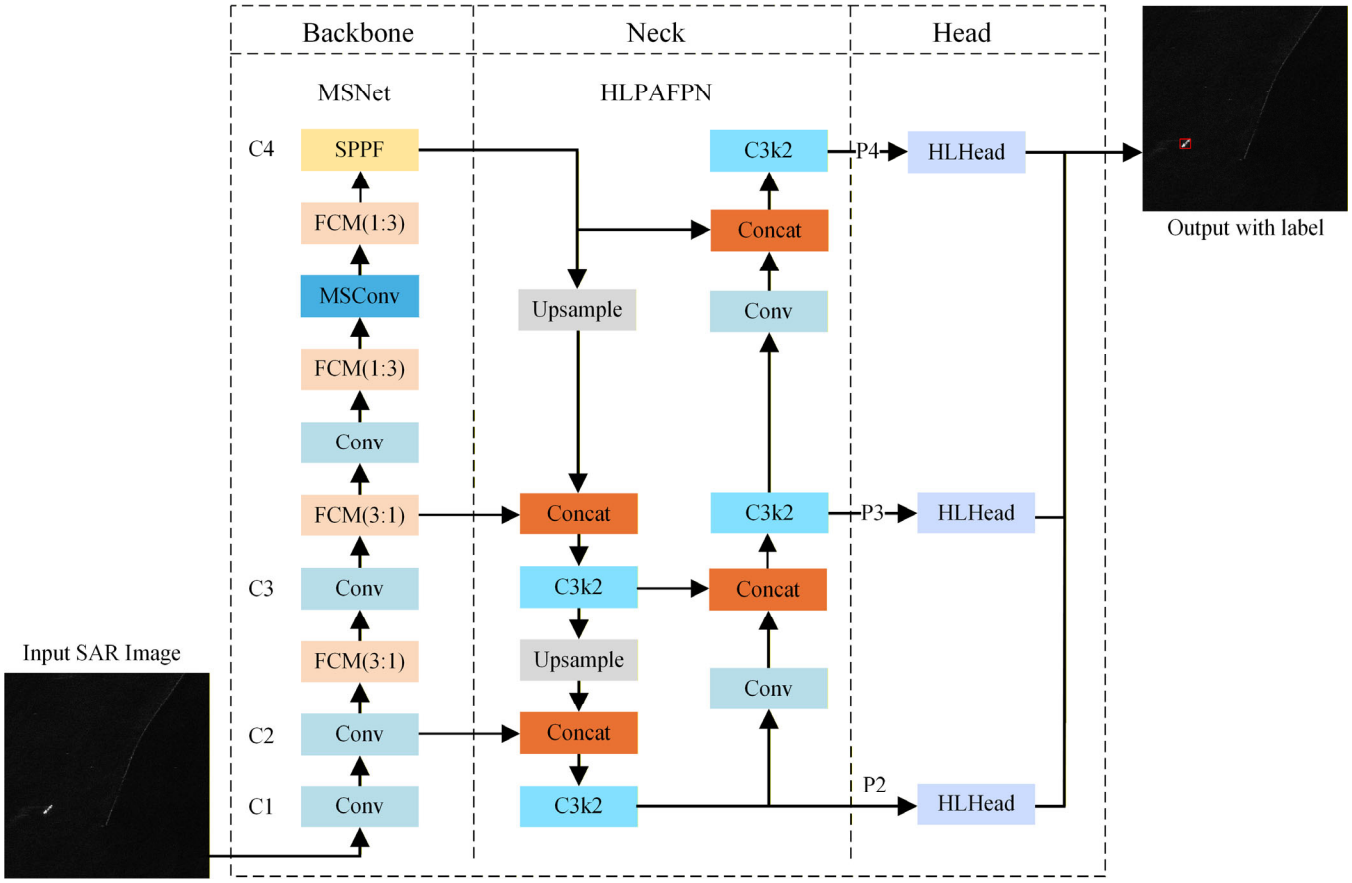


Figure 2. The overall structure of High-Lightweight Net (HLNet), showing the complete pipeline from SAR image input to detection output.

First, in MSNet, we introduced the feature complementary mapping module. The FCM module alleviates spatial information loss in SAR images during downsampling via channel complementarity and spatial interaction, thereby enhancing the fusion of features and semantic information for small ships. Furthermore, the MSConv module replaced 3×3 convolutions in deeper network layers. By leveraging multi-core parallel convolutions, MSConv attenuates high-frequency noise while amplifying response weights for medium-to-large ships.

Second, we refined the feature fusion path of the conventional FPN + PAN architecture and restructured the neck network. Through channel adaptation, the neck was aligned with the spatial-semantic features output by the backbone. Additionally, redundant large-scale layers were eliminated. These modifications collectively enhance feature fusion efficiency and optimize computational efficiency.

Finally, to address the high computational cost of the original detection head module, we designed an efficient, lightweight head by integrating grouped convolutions and DFL.

3.2. Multi-Scale Net (MSNet)

Standard convolutions struggle to simultaneously attenuate coherent speckle noise and extract complete features of large ships in SAR images. To address this challenge, we designed a task-specific backbone architecture called MSNet, with its complete structure depicted in Figure 3.

MSNet performs initial downsampling and feature extraction via a series of feedforward convolutions. At the early and intermediate layers of the network, multiple FCM modules are embedded. The FCM module effectively encodes and transmits detailed

spatial positional information from early to deeper levels by employing channel splitting and complementary mapping. This design significantly alleviates the loss of spatial information for small targets as the network deepens. In the deeper layers of the network, the MSConv module is incorporated to enhance feature discriminability. The MSConv module parallelizes convolutional kernels of different sizes (3×3, 5×5, 7×7) to capture multi-scale contextual information. Meanwhile, it integrates a channel attention mechanism that adapts through channel weight recalibration, enabling the network to focus on more discriminative ship features. This design effectively suppresses complex coherent speckle noise interference in SAR images.

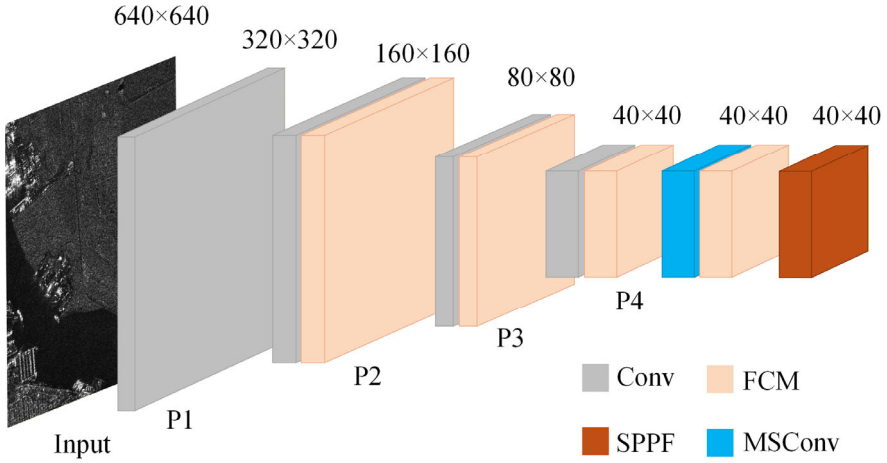


Figure 3. The overall structure of Multi-Scale Net (MSNet).

The workflow of MSNet is as follows. First, the input image undergoes two convolutional operations with a stride of 2, reducing its resolution to one-quarter and extracting preliminary features. The resulting features then pass through an FCM module. At this stage, the FCM module refines the features via its complementary mapping mechanism, actively preserving and enhancing spatial detail information to support the detection of small ships in later stages. Subsequently, the feature maps are downsampled to 1/8 and 1/16 of the original resolutions. After each downsampling step, an FCM module is inserted. This staged deployment ensures that as the network depth increases, semantic information is progressively enriched, while the precise spatial details from early layers are continuously retained and fused. This process effectively alleviates the vanishing of small targets in the feature maps. Following this, the network proceeds to the deep feature extraction stage, where conventional convolutions are replaced by the MSConv module. The core innovation of this module lies in its multi-kernel parallel architecture and attention-augmented mechanism. The module employs convolutional kernels of multiple sizes for parallel feature extraction, accommodating the wide range of ship scales in complex scenarios. Its integrated attention mechanism allows the network to autonomously learn channel-wise importance, thereby focusing on the features most relevant to ships and improving feature discriminability. Finally, the processed features pass through an SPPF module with optimized channel counts before being fed into the HLPAFPN for local and global feature fusion.

3.2.1. Feature Complementary Mapping (FCM)

In deep neural networks, the spatial information of small objects is highly susceptible to loss. To alleviate this, the FCM module employs a "divide-and-conquer" strategy. Specifically, the input features are split channel-wise into two parallel branches: a spatial detail branch (x^1) and a semantic context branch (x^2).

The spatial detail branch (x^1) employs multi-layer convolutions with small receptive fields to focus on local detail extraction, thereby enhancing the capture of fine-grained structures such as ship edges and corners, producing channel-enhanced features denoted by $x^{channel}$. Conversely, the semantic context branch (x^2) utilizes lightweight 1×1 convolutions to preserve global semantic information, facilitating holistic contour modeling for targets across various scales and generating spatially modulated features denoted by $x^{spatial}$. Subsequently, x^1 is passed through a spatial attention module composed of depthwise convolution and adaptive average pooling to enhance spatial responses in target regions, whereas x^2 is processed by a channel attention module built with pointwise convolution and batch normalization, which recalibrates discriminative channel features to better distinguish ships from the background.

In the proposed FCM, the relative contribution ratios between the two branches are manually specified rather than dynamically determined. As illustrated in Figure 2, the first two modules adopt a fixed ratio of 3:1 to emphasize spatial detail extraction, while the subsequent two modules use a ratio of 1:3 to strengthen semantic context aggregation. These ratios are determined based on preliminary experiments and ablation studies, aiming to balance fine-grained structural information and global semantic representation.

After the attention mechanisms are applied, element-wise multiplication is performed, whereby the x^1 branch is enhanced with refined spatial information from $x^{spatial}$, and the x^2 branch is reinforced with channel-wise information from $x^{channel}$. Finally, the enhanced features from both branches are integrated via element-wise addition, generating an output feature map with the same channel dimension as the input. This process yields a well-balanced representation that preserves both detailed recognition and semantic integrity for subsequent feature fusion stages. The overall architecture of the FCM module is depicted in Figure 4.

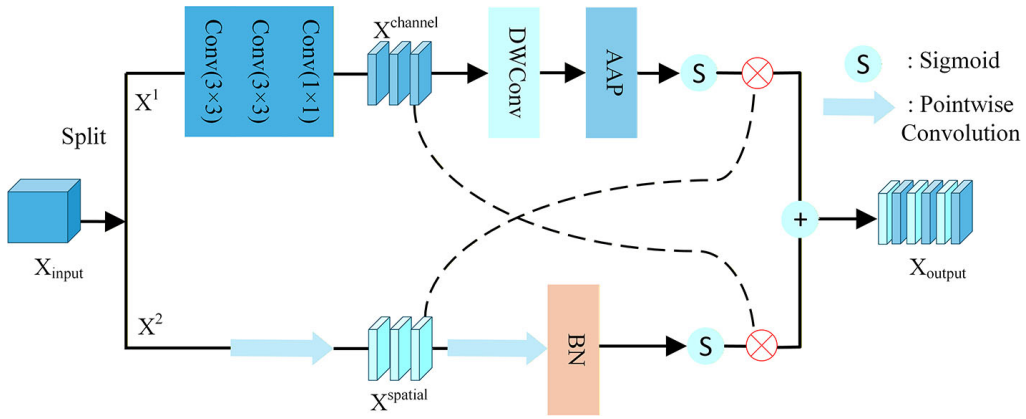


Figure 4. The overall architecture of the Feature Complementary Mapping (FCM) module.

3.2.2. Multi-Scale Convolution (MSConv)

Drawing inspiration from the Inception architecture and tailored for SAR characteristics, the MSConv module parallelizes three sets of depthwise separable convolutions (DWConv) with kernel sizes of 3×3, 5×5, and 7×7, as illustrated in Figure 5a. The 3×3 small kernel focuses on capturing local textures while attenuating high-frequency speckle noise. The larger 5×5 and 7×7 kernels expand the receptive field to ensure the integrity of large ship features, effectively distinguishing them from surrounding backgrounds such as wave clutter and coastlines, thereby delineating their boundaries more precisely. Then, features from each branch are integrated across channels via pointwise convolution and concatenated, achieving a multi-scale feature representation.

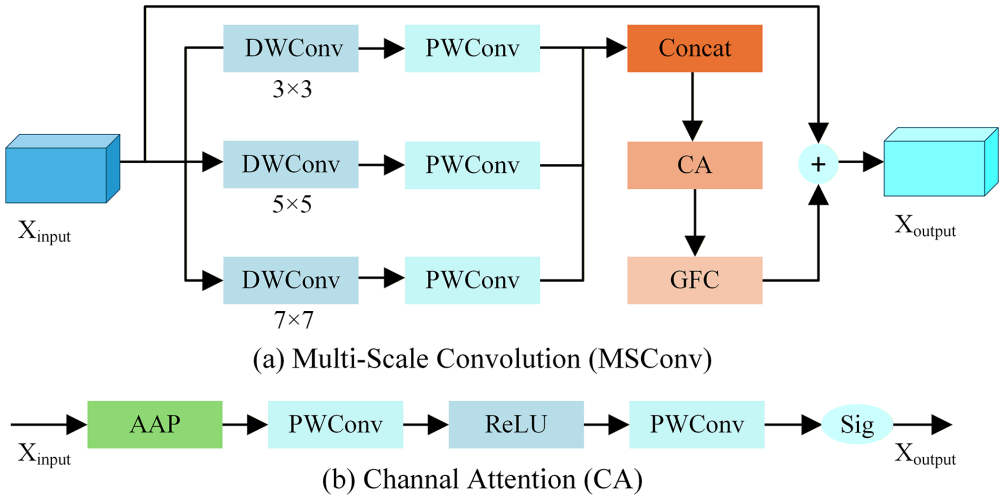


Figure 5. The workflow structure of the MSConv. (**a**) The structure of MSConv. (**b**) The structure of CA.

Following this parallel feature concatenation, a lightweight Channel Attention (CA) module is incorporated. It autonomously emphasizes feature channels relevant to ships while suppressing background noise responses. Specifically, it first obtains a global channel descriptor via adaptive average pooling. Subsequently, channel-wise weight coefficients are generated through a lightweight bottleneck structure consisting of two pointwise convolutions separated by a ReLU activation, followed by a sigmoid function. This process enables the autonomous recalibration of critical feature channels. Next, a Grouped Fusion Convolution (GFC) with a grouping factor of 8 is applied. The GFC promotes interaction across different feature subspaces while maintaining parameter efficiency. Finally, a residual connection is established between the output of the GFC and the initial input features to the MSConv module. This design not only preserves the integrity of the original information and alleviates gradient vanishing but also allows the network to focus on learning residual components that are more discriminative for the final detection task.

3.2.3. High-Lightweight Feature Pyramid (HLPAFPN)

While the traditional FPN + PAN architecture demonstrates strong performance in general object detection, its design is redundant for SAR ship detection due to the relatively homogeneous backgrounds and the dominance of small targets in SAR imagery. Accordingly, HLPAFPN implements key pruning and enhancement operations on the standard architecture, as illustrated in Figure 6.

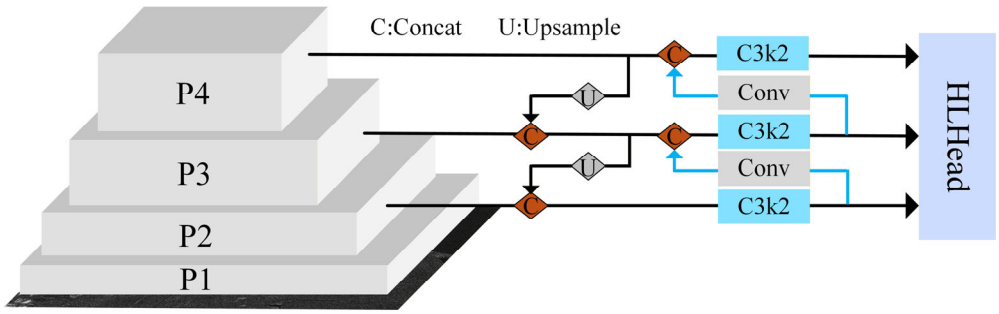


Figure 6. The overall structure of High-Lightweight Feature Pyramid (HLPAFPN).

First, to enhance micro-scale perception, an additional high-resolution feature fusion path (the P2 layer) was introduced. This path directly injects fine-grained information

from the early layers of the backbone network into the feature pyramid. Specifically, deep features from the backbone are upsampled and concatenated with the P3-layer features. This composite feature map is further upsampled to the highest resolution (P2 layer), constructing a dedicated enhancement layer for small object detection. This design fully leverages the detailed low-level features from MSNet, thereby significantly improving the detection capability for small-scale ships in SAR images.

Second, given that SAR ship targets rarely appear at extremely large scales, the fusion at the P5 level was eliminated, and the maximum downsampling ratio was restricted to 16× to avoid redundant computations.

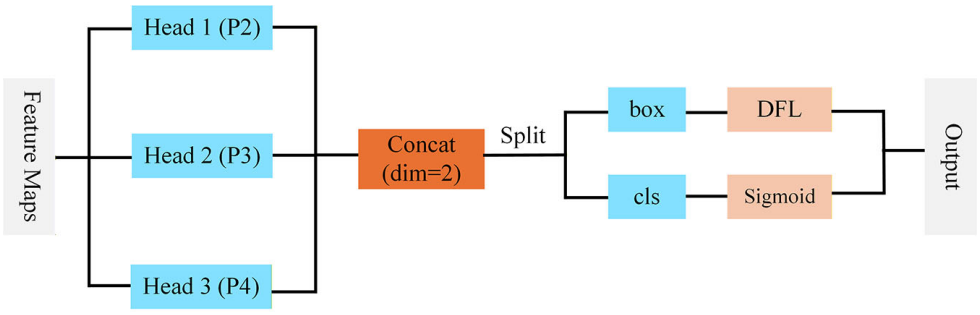
Finally, experimental validation showed that reducing the channel count in the fusion layers by 50% not only maintained but slightly improved accuracy by mitigating overfitting risks, while simultaneously achieving a substantial reduction in model parameters.

The workflow of HLPAFPN is as follows. A feature map of dimensions $40 \times 40 \times 512$ acts as the input. This map is first upsampled and fused with the P3-layer features from the backbone network, yielding an intermediate feature map of size 80×80. The resulting feature map is then further upsampled to 160×160 and concatenated with the spatially detailed P2-layer features, forming an enhanced layer optimized for ship detection. Following this, two stride-2 convolutional downsampling operations are applied. Through these operations, the high-resolution details from the enhanced layer are progressively integrated, first with the P3-layer features in the neck and then with the deeper features from the backbone. This hierarchical integration ensures that the output feature maps preserve crucial positional information from the early layers.

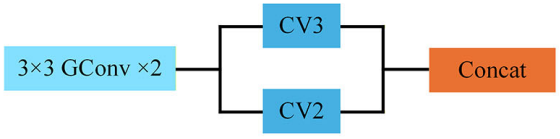
Overall, this design systematically aggregates multi-scale information spanning from spatial details to deep semantics. It achieves an effective balance between detection accuracy and inference speed, while also facilitating a significant reduction in the number of parameters.

3.3. High-Lightweight Head (HLHead)

The proposed HLHead adopts a decoupled detection head to improve the accuracy and robustness of multi-scale ship detection in SAR imagery, as illustrated in Figure 7. It consists of three parallel detection heads corresponding to feature maps P2, P3, and P4 from the HLPAFPN neck, enabling effective detection of ships across different scales.



(a) Overall pipeline of the High-Lightweight Head (HLHead).



(b) Detailed internal structure of a single detection head.

Figure 7. The HLHead module architecture. In (**b**), cv2 is a 1 × 1 convolution projecting to 4 × reg_max channels for box regression, cv3 is a 1 × 1 convolution projecting to nc channels for classification, followed by channel-wise concatenation.

Within each detection head (Figure 7b), the input feature map is first processed by a stem module composed of two consecutive 3 × 3 group convolutions. Compared with standard convolutions, group convolution significantly reduces computational cost and parameter overhead while preserving sufficient inter-channel feature representation, which is beneficial for efficient inference in real-time SAR ship detection scenarios.

The refined features are then fed into two parallel 1 × 1 convolutional branches. The regression branch projects the features into 4 × reg_max channels for bounding box prediction (where reg_max is the number of discrete bins per boundary direction in the DFL formulation), while the classification branch outputs nc channels for category prediction. The outputs of both branches are concatenated along the channel dimension to form per-scale predictions. Subsequently, predictions from all scales are concatenated along the spatial dimension to form a unified prediction tensor, which is then split into box and classification branches for joint decoding. The classification branch applies a sigmoid activation to obtain class probabilities, whereas the regression branch adopts the DFL mechanism for bounding box localization.

The use of DFL is motivated by the intrinsic characteristics of SAR imagery. Due to speckle noise and low contrast between ship boundaries and sea clutter, target edges in SAR images are often ambiguous, particularly for small-scale ships. Under such conditions, direct coordinate regression with conventional losses (e.g., Smooth L1 or IoU-based losses) tends to be unstable and sensitive to noise, leading to increased localization errors. DFL mitigates this issue by reformulating bounding box regression as a discrete probability distribution learning problem, enabling the model to explicitly capture localization uncertainty.

Specifically, instead of predicting a single continuous coordinate, DFL predicts a discrete probability distribution over a set of predefined boundary positions. The final coordinate is obtained as the expectation of the predicted distribution, as formulated in Equation (1). During training, the DFL (Equation (2)) focuses on the two discrete positions that enclose the ground-truth value, ensuring accurate localization while maintaining robustness to boundary ambiguity. This distribution-based modeling is particularly effective under the low signal-to-noise ratio conditions commonly encountered in SAR imagery.

$$\hat{y} = \sum_{i=0}^{n} P(y_i) y_i \tag{1}$$

where y_i denotes the discrete boundary position and $P(y_i)$ is the predicted probability, and the final coordinate is obtained as the expectation of the learned distribution.

$$L_{DFL} = -\left((y_{i+1} - y)\log(P(y_i)) + (y - y_i)\log(P(y_{i+1}))\right) \tag{2}$$

The DFL assigns supervision only to the two discrete positions that enclose the ground-truth value, encouraging accurate localization while preserving uncertainty modeling.

Finally, the DFL-decoded bounding boxes are combined with the sigmoid-activated class probabilities to generate the final detection results. In addition, in HLHead, setting the group number g to match the number of DFL intervals makes the grouped convolution design naturally compatible with distribution-based regression, promoting efficient feature extraction for distribution-based bounding box regression.

4. Experiments and Results

4.1. Environment and Configuration

The hardware and software configuration used in the experiments is summarized below: The model was trained on a workstation equipped with an AMD Ryzen 9 7900X 12-core processor (Advanced Micro Devices, Santa Clara, CA, USA) and an NVIDIA GeForce RTX 3090 GPU (NVIDIA Corporation, Santa Clara, CA, USA), running Windows

11 Pro (Microsoft Corporation, Redmond, WA, USA). The software environment was established upon the PyTorch 2.7.1 deep learning framework (Meta Platforms, Inc., Menlo Park, CA, USA) with CUDA 11.8 (NVIDIA Corporation, Santa Clara, CA, USA). Training employed the stochastic gradient descent (SGD) optimizer with the initial learning rate, momentum, and weight decay set to 0.01, 0.937, and 0.0005, respectively. Comprehensive hyperparameter configurations are summarized in Table 1.

Table 1. Specific Hyperparameter Settings for the Model.

Hyperparameters	Value
Epochs	200
Image Size	640 × 640
Batch Size	36
Workers	4
Pretrained	False

4.2. Dataset Description

To evaluate the effectiveness of HLNet, experiments were conducted on three datasets: SSDD, HRSID, and the Complex SAR Images Dataset (CSID). For SSDD and HRSID, the data were split into training, validation, and test sets with a ratio of 7:1:2. Both SSDD and HRSID are widely used public benchmarks for ship detection in the SAR imaging domain. SSDD contains a higher proportion of small-scale targets, whereas HRSID exhibits larger-scale variations and more complex scenes. Additionally, the CSID proposed in this paper is a challenging composite subset constructed from SSDD and HRSID, specifically designed for ship detection in complex background scenarios.

4.2.1. SSDD

The SSDD dataset, released in 2017, is one of the earliest and most widely used public benchmark datasets in the field of SAR ship detection. The dataset was collected from multiple satellites, including Sentinel-1, RADARSAT-2, and TerraSAR, and comprises 1160 SAR images containing 2456 ship targets. SSDD images were acquired in diverse coastal and offshore scenarios, with spatial resolutions ranging from 1 to 15 m and polarization modes including HH, HV, VV, and VH. It is widely used as a reference benchmark for evaluating the robustness and generalization capability of ship detection methods.

4.2.2. HRSID

HRSID is an open benchmark dataset constructed for ship detection and instance segmentation in SAR imagery, introduced by Su et al. [31]. It aggregates SAR imagery from Sentinel-1 and TerraSAR satellites, processed into 5604 image tiles of size 800 × 800, with pixel-level segmentation masks and bounding box annotations for 16,951 ship instances. The dataset encompasses diverse complex environments, ranging from open seas to coastal areas and ports, covering a wider range of ship target scales and enabling effective evaluation of algorithm performance under challenging conditions.

4.2.3. CSID

A joint subset, termed CSID, was manually selected from the SSDD and HRSID datasets, comprising 379 SAR images featuring complex backgrounds. This dataset contains 2000 ship targets under extremely complex background conditions, featuring dense vessel clusters and varying scales, effectively testing the robustness of ship detection methods in complex scenarios.

Table 2 summarizes the statistical characteristics of ship targets in the SSDD, HRSID, and the proposed CSID subset, including the number of images and ship instances, image

spatial resolution, the average and standard deviation of ship width, height, and area and the average target density. As shown in Table 2, SSDD and HRSID exhibit comparable average ship widths, heights, and areas, indicating that both datasets are dominated by small-scale targets. However, HRSID presents a significantly larger standard deviation of ship area, reflecting greater dispersion in target size and a wider range of ship scales.

Table 2. Statistical comparison of ship targets in SSDD, HRSID, and the proposed CSID subset. CSID exhibits significantly higher target density and much more severe scale imbalance.

Dataset	Images	Ships	Resolution (m)	Width (px)		Height (px)		Area (px^2)		Avg. Density
				Avg.	Std.	Avg.	Std.	Avg.	Std.	
SSDD	1160	2587	1–15	34	35	38	33	1678	3294	2.23
HRSID	5604	16,950	0.5, 1, 3	37	33	33	30	1809	8236	3.02
CSID	379	2000	1–15	44	54	37	39	3424	18,511	5.28

In contrast, CSID demonstrates markedly different characteristics, with substantially higher target density and much larger size dispersion. Specifically, CSID contains on average more than five ships per image, and its standard deviation of ship area exceeds five times the mean value, revealing an extremely skewed scale distribution. These statistics confirm that CSID represents a highly challenging subset with dense and multi-scale ship distributions, making it suitable for stress-testing the robustness and generalization ability of HLNet under complex maritime scenarios.

As shown in Figure 8, the SSDD dataset predominantly features small- to medium-sized vessels, with widths and heights mainly concentrated between 0 and 150 pixels, reflecting a focus on compact targets. In contrast, the HRSID dataset includes ship targets across a considerably broader scale range, incorporating numerous large vessels exceeding 400 pixels in both width and height dimensions. The CSID dataset, derived as a challenging subset from SSDD and HRSID (selected based on criteria such as high occlusion, dense clustering, and extreme scale variations), exhibits a distribution similar to HRSID but with an emphasis on more difficult instances, including ship targets with higher density and a wider scale range. These differing scale emphases across the SSDD and HRSID datasets provide complementary characteristics essential for training and evaluating HLNet's multi-scale detection capability. Notably, as a challenging test subset, CSID is used to verify the model's generalization ability on difficult samples. In our subsequent experiments, the pre-trained weights on HRSID are directly adopted for transfer testing.

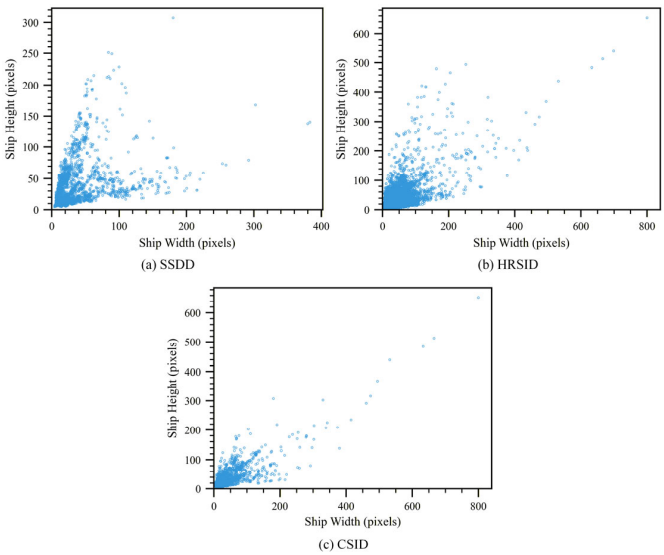


Figure 8. Distribution of ship sizes in the (**a**) SSDD, (**b**) HRSID, and (**c**) CSID datasets.

4.3. Evaluation Metrics

To objectively evaluate model performance, commonly used object detection metrics are adopted, including precision, recall, F1 score, and mean average precision (mAP). Additionally, to quantify model complexity and real-time performance, three additional indicators are considered: the number of model parameters, floating-point operations (FLOPs), and frames per second (FPS). The corresponding calculation formulas are defined in Equations (3)–(7).

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{5}$$

$$AP = \int_0^1 P(R)dR \tag{6}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{7}$$

In SAR ship detection, TP denotes correctly detected ship targets, FP represents false alarms including non-ship objects or duplicate detections, and FN denotes missed ground-truth ships. Precision (P) and recall (R) respectively measure detection accuracy and completeness, while F1 harmonizes their trade-off. AP evaluates category-wise detection performance via the precision–recall curve, and mAP is the mean AP over all categories. Model efficiency is assessed in terms of parameter count, FLOPs, and FPS, reflecting spatial complexity, computational cost, and real-time capability.

4.4. Ablation Experiments

4.4.1. Ablation Experiments on Publicly Available Datasets

To evaluate the contribution of each module in HLNet, ablation experiments were conducted on the SSDD and HRSID datasets using four configurations. YOLOv11 was adopted as the baseline, upon which MSNet, HLPAFPN, and HLHead were progressively integrated to assess their effects on feature extraction, feature fusion, and model efficiency, respectively.

The ablation results are summarized in Table 3. On the SSDD dataset, the proposed method achieved mAP_{50} and mAP_{75} of 98.3% and 88.8%, improving over the baseline by 1.2% and 2.8%, respectively, with a 1.6% gain in F1 score. On the HRSID dataset, mAP_{50} and mAP_{75} reached 91.7% (+2.0%) and 78.0% (+6.2%), while the F1 score increased to 0.875 from 0.836.

Table 3. Ablation experiments on the SSDD and HRSID.

MSNet	HLPAFPN	HLHead	SSDD			HRSID			Params (M)	FLOPs (G)
			mAP_{50} (%)	mAP_{75} (%)	F1	mAP_{50} (%)	mAP_{75} (%)	F1		
			97.1	86	0.937	88.7	71.8	0.836	2.58	6.3
√			97.6	87.5	0.941	90.6	75.7	0.855	2.67	10.3
√	√		97.7	87.8	0.952	91.8	77.2	0.873	0.85	9
√	√	√	98.3	88.8	0.953	91.7	78	0.875	0.66	5.4

The introduction of MSNet enhanced feature extraction capability at the cost of increased computational complexity. With the integration of HLPAFPN, the parameter count was reduced to approximately one-third of the original without compromising accuracy, demonstrating the effectiveness of SAR-oriented channel pruning. After incorporating HLHead, the model achieved an extremely lightweight design with only 0.66 M parameters,

corresponding to a 72% reduction in parameters and a 14% reduction in FLOPs compared to the baseline.

Notably, on the HRSID dataset, HLHead led to a slight decrease (≈0.1%) in mAP_{50} while increasing mAP_{75} by 0.8%. This phenomenon can be attributed to the discretized nature of DFL-based regression, which is less sensitive at lower IoU thresholds for large objects but exhibits greater robustness at higher IoU thresholds.

To provide a more intuitive understanding of the feature regions attended by the model, LayerCAM [44] heatmaps were employed for visual ablation analysis. As shown in Figure 9, compared to the baseline model in Figure 9a, the results from HLNet in Figure 9d demonstrated greater focus on the vessel target, with notably improved background suppression. While the baseline model can roughly activate on the hull region in the SAR image, it fails to consistently concentrate attention on the vessel body, and the corresponding heatmap exhibits noticeable edge diffusion.

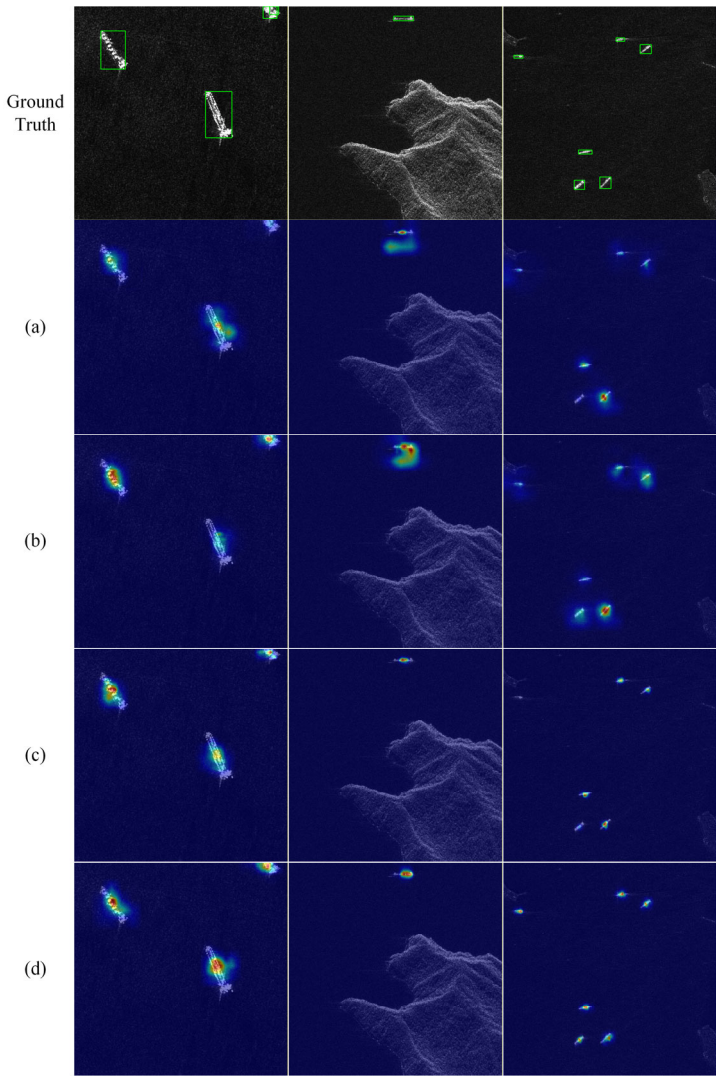


Figure 9. The results of the visual ablation experiment: (**a**) shows the baseline output; (**b**) shows the baseline-MSNet output; (**c**) shows the baseline-MSNet-HLPAFPN output; (**d**) shows the baseline-MSNet-HLPAFPN-HLHead (HLNet) output.

This diffusion phenomenon indicates insufficient learning of ship scattering characteristics and a tendency to misinterpret weak edge signals as target responses. With the introduction of MSNet, HLPAFPN, and HLHead, the diffusion at the heatmap edges gradually converged toward the vessel area, indicating the models increasingly focused on

the hull region within the image. The experimental results demonstrated that the proposed method could effectively extract ship targets while attenuating background noise, thereby enhancing the overall detection performance.

4.4.2. Hyperparameter Analysis Experiments

To analyze the effect of the input feature splitting ratio on FCM performance, comparative experiments were conducted on the SSDD dataset using fixed and dynamic channel allocation strategies. As shown in Table 4, the proposed dynamic scheduling scheme (transitioning from 3:1 to 1:3) achieved the best performance, yielding the highest $mAP_{50–95}$ (72.5%) and FPS (354). In contrast, the fixed 3:1 ratio excessively compressed the spatial attention branch, limiting fine-grained spatial representation and resulting in the lowest $mAP_{50–95}$ (71.8%) and higher computational cost.

Table 4. Experiment on feature partition ratio in FCM. Bolded numbers indicate optimal indicators.

Partition Ratio	$mAP_{50–95}$ (%)	Params (M)	FLOPs (G)	FPS
1:3	72.3	**0.61**	**4.4**	330
3:1	71.8	0.95	6.3	337
3:1–1:3	**72.5**	0.66	5.4	**354**

Conversely, the 1:3 scheme reduced parameters and FLOPs but constrained representational capacity, leading to suboptimal accuracy and a 24 FPS reduction compared to the dynamic strategy. This apparent discrepancy—lower FLOPs but reduced FPS—can be attributed to hardware efficiency and parallelism considerations. Specifically, allocating fewer channels to the spatial attention branch reduces the computation per channel, which may underutilize the GPU's parallel processing capability. Additionally, the smaller channel width can increase memory access frequency for intermediate feature maps, raising the Memory Access Cost (MAC) and limiting throughput. In contrast, the dynamic scheduling strategy adjusts the partitioning ratio during inference, maintaining a balance between semantic and spatial information while achieving higher computational density and better parallel efficiency, thereby yielding both higher accuracy and faster inference. These results demonstrate that the proposed dynamic strategy effectively balances representational capacity and hardware utilization, achieving an optimal trade-off between accuracy and efficiency.

Ablation experiments were performed with g = 8, 16, and 32 to assess the effect of group count in HLHead and the necessity of aligning g with DFL intervals. The results in Table 5 show that, compared with g = 8, g = 16 improves $mAP_{50–95}$ by 0.6% while reducing parameters and FLOPs by 8.3% and 12.9%, respectively, and achieving the fastest inference speed (354 FPS).

Table 5. Experiment on the number of groups in grouped convolution. Bold numbers indicate the best indicators.

Number of Groups	$mAP_{50–95}$ (%)	Params (M)	FLOPs (G)	FPS
g = 8	71.9	0.72	6.2	300
g = 16	**72.5**	0.66	5.4	**354**
g = 32	71.4	**0.63**	**5**	322

Although g = 32 has fewer parameters and lower FLOPs, its inference is slower (322 FPS). This counterintuitive result may be due to the increased group number, which reduces the number of channels per group, lowering per-group computational density and underutilizing GPU parallelism. As a consequence, the execution overhead associated

with kernel launch and synchronization becomes more pronounced, negatively affecting overall inference throughput. Smaller groups also lead to more fragmented memory access, decreasing cache hit rates and increasing memory latency.

By contrast, g = 16 strikes a favorable balance: each group is sufficiently large to maintain high computational density and efficient cache utilization, maximizing parallel throughput. When g is reduced to 8, inter-group communication becomes more intensive, slightly increasing FLOPs and parameters while reducing parallel efficiency, resulting in lower FPS (300). Overall, aligning the group number (g = 16) with the DFL interval count provides an optimal trade-off between feature granularity, detection accuracy, and inference efficiency, while effectively leveraging hardware resources.

4.5. Comparison Experiments

4.5.1. Comparative Experiments on HRSID

We selected several state-of-the-art object detection algorithms, including the widely adopted one-stage detector YOLOv8, the lightweight and high-performance YOLOv9-Tiny, and the real-time detector RT-DETR based on an end-to-end Transformer architecture. The baseline model YOLO11 was also included for comparison. Comparative experiments were conducted on the HRSID and SSDD datasets to validate the effectiveness of HLNet. To ensure a fair comparison, consistent parameter configurations were maintained for all detectors during training.

The comparative experimental results on HRSID are presented in Table 6. Across metrics including detection performance, model compactness, and real-time capability, HLNet comprehensively outperformed YOLOv8, YOLOv9-Tiny, YOLO11, and RT-DETR. HLNet achieved an F1 score of 0.875 and mAP_{50} of 91.7%, with an inference speed of 405 FPS. Meanwhile, its model parameters and FLOPs were only 0.6 million and 5.4 GFLOPs, respectively, further validating the effectiveness of HLNet.

Table 6. Comparison with classical object detection algorithms on the HRSID dataset. Bold numbers indicate the best indicators.

Method	F1 (%)	mAP_{50} (%)	Params (M)	FLOPs (G)	FPS
YOLOv8	0.857	89.3	3	8.1	363
YOLOv9-tiny	0.860	89.3	2.6	10.7	277
YOLO11	0.852	88.7	2.5	6.4	395
RT-DETR	0.861	90.1	20.1	60.1	47
HLNet	**0.875**	**91.7**	**0.6**	**5.4**	**405**

Compared with YOLOv8 and YOLOv9-Tiny, HLNet achieves consistent gains in F1 score (up to +1.8%) and mAP_{50} (+2.4%), while requiring only about 20% of YOLOv8's parameters and reducing parameters and GFLOPs by approximately 76% and 50% relative to YOLOv9-Tiny. In comparison with RT-DETR, HLNet delivers superior detection accuracy and faster inference with substantially lower computational complexity.

In addition, relative to the baseline YOLO11, HLNet improves F1 and mAP_{50} by 2.3% and 3.0%, respectively, while reducing parameters by 80% and FLOPs by 15.62%. Experimental results demonstrated that HLNet achieves enhanced detection accuracy and inference speed, with the fewest parameters and computational cost among the compared models.

4.5.2. Comparative Experiments on SSDD

The comparative experimental results on SSDD (Table 7) demonstrated that the proposed HLNet achieved the highest F1 score and mAP_{50}, with values of 0.953 and 98.3%, respectively. Compared to YOLOv8, YOLOv9-Tiny, RT-DETR, and YOLO11, HLNet's

mAP_{50} was 1.4%, 0.6%, 1.2%, and 4.1% higher, respectively. Among the detectors used in these comparative experiments, YOLO11 had the smallest parameter count at 2.5 million, which is approximately four times larger than that of HLNet. Additionally, HLNet achieved an inference speed of 243 FPS, demonstrating superior lightweight characteristics and real-time performance.

Table 7. Comparison with classical object detection algorithms on the SSDD dataset. Bold numbers indicate the best indicators.

Method	F1	mAP_{50} (%)	Params (M)	FLOPs (G)	FPS
YOLOv8	0.928	96.9	3	8.1	180
YOLOv9-Tiny	0.934	97.7	2.6	10.7	125
YOLO11	0.937	97.1	2.5	6.4	209
RT-DETR	0.913	94.2	20.1	60.1	68
HLNet	**0.953**	**98.3**	**0.6**	**5.4**	**243**

4.5.3. Comparative Experiments on CSID

To validate detection performance in complex scenarios, the CSID dataset—characterized by intricate backgrounds and diverse target scales—was adopted as the evaluation benchmark. CSID contains 379 SAR images with 2000 ship instances, exhibiting higher background complexity and target density than HRSID and SSDD, providing a more challenging test. Models trained on HRSID were directly evaluated on CSID. As shown in Table 8, all comparison methods suffer notable performance degradation, highlighting the difficulty of SAR ship detection in complex environments. In contrast, HLNet consistently achieves superior accuracy, lower parameter count, and faster inference speed, demonstrating strong performance, efficiency, and generalization under challenging conditions.

Table 8. Comparison with classical object detection algorithms on the CSID dataset. Bold numbers indicate the best indicators.

Method	F1	mAP_{50} (%)	Params (M)	FLOPs (G)	FPS
YOLOv8	0.722	72.2	3	8.1	102
YOLOv9-Tiny	0.675	70.7	2.6	10.7	72
YOLO11	0.713	71.6	2.5	6.4	109
RT-DETR	0.653	68.4	20.1	60.1	37
HLNet	**0.758**	**75.9**	**0.6**	**5.4**	**112**

4.5.4. Visualization Results of the Comparative Experiments

To evaluate HLNet's performance across diverse backgrounds, three challenging images were selected for qualitative comparison. Figure 10a features a complex terrestrial background with urban coastlines and river confluences, which tested the detectors' capabilities in intricate land environments. YOLOv8, YOLOv9-Tiny, and RT-DETR all exhibited significant false positives and false negatives. Additionally, YOLO11 missed detecting one ship, while only the proposed HLNet closely matched the ground truth. Figure 10b features densely packed ships of varying scales, challenging the distinguishing capability and multi-scale detection capability of the model. The results showed that YOLOv8 and YOLOv9-Tiny struggled with dense object scenarios, exhibiting severe false positives and negatives. In contrast, YOLO11 performed slightly better but still produced two detection errors. Only one ship was missed by both RT-DETR and our model. This might be attributed to information loss caused by downsampling, which may have led to bounding box merging when the model dealt with small vessels that were completely adjacent to each other. Figure 10c depicts scenarios such as coastlines and narrow waterways, where the background and vessels are even more difficult to distinguish with the naked eye, thereby

demanding comprehensive capabilities from the detector. Under this more challenging interference scenario, both YOLOv9-Tiny and RT-DETR exhibited significant performance degradation, with each model detecting only partial objects. YOLOv8 and YOLO11 performed slightly better but still exhibited false negatives and false positives, whereas HLNet produced no errors. In summary, the proposed HLNet effectively addressed the challenge of multi-scale maritime object detection under complex backgrounds.

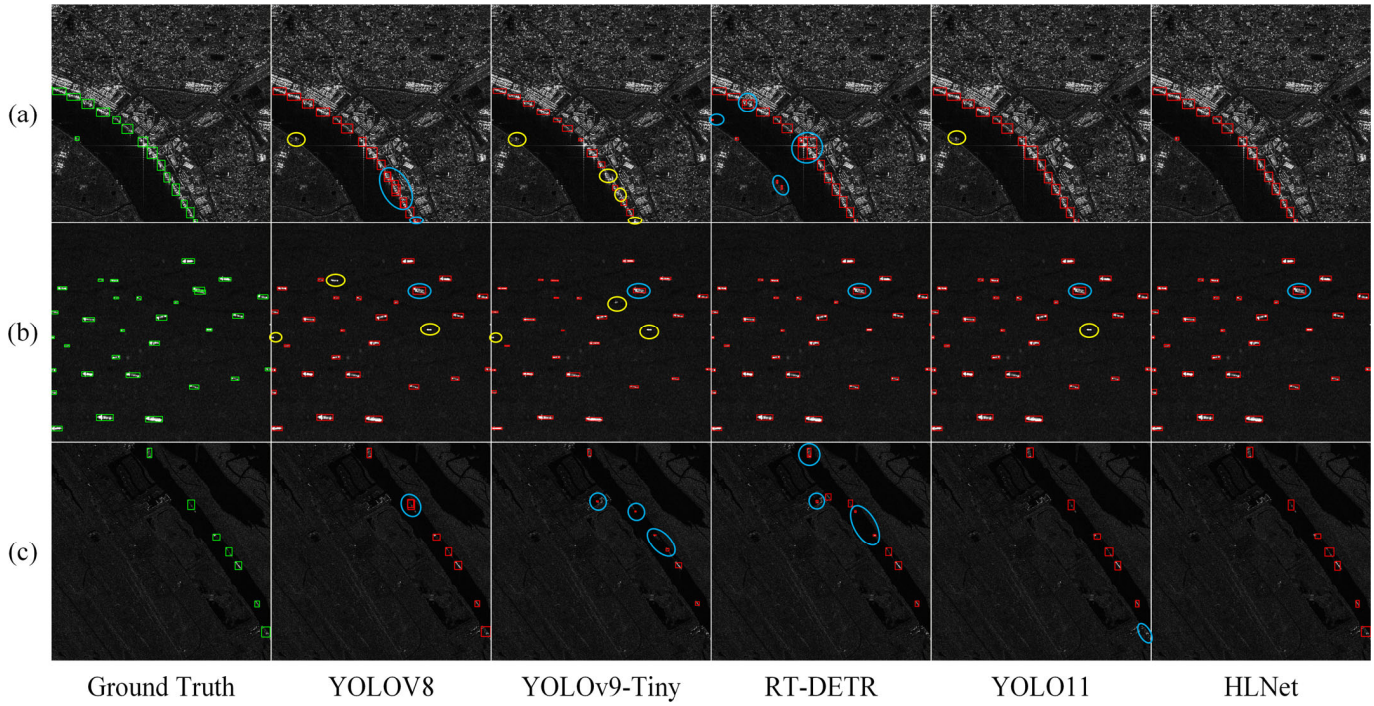


Figure 10. Comparative experiment of SAR sample images. The yellow circle represents the missed detection area, and the blue circle represents the false detection area.

4.5.5. Comparison with Other SAR Ship Detection Models

To further validate the performance of HLNet, several models developed in recent years for SAR-based ship detection were selected for comparison. Since most SAR ship detectors do not release their source code, only a limited comparison could be conducted based on the results reported in the original publications.

In the SSDD experimental results (Table 9), HLNet demonstrated outstanding overall performance. Its mAP_{50} reached 98.3%, outperforming multiple SAR ship detection models proposed in recent years. Although EsarDet achieved a slightly higher F1 score of 0.96, HLNet's mAP_{50} was 0.4% higher while reducing the number of model parameters by approximately 90% and FLOPs by 28%. Comprehensive analysis indicates that HLNet delivered superior overall performance. Additionally, in terms of computational complexity, AC-YOLO shared the same 5.4 GFLOPs as HLNet. However, HLNet achieved a 1.6% improvement in mAP_{50} while reducing the number of parameters by approximately 67%. In summary, HLNet demonstrated significant advantages in both detection accuracy and model compactness. These results collectively demonstrate the effectiveness and deployment potential of HLNet for SAR-based ship detection.

In the comparative experiments on the HRSID dataset, HLNet still achieved the best performance in ship detection. Its F1 score reached 0.875 with a mAP_{50} of 91.7%. Although Chen's method achieved the same mAP_{50} as HLNet, its computational load was 17 times higher, indicating that accuracy gains came at the expense of computational efficiency. Compared with BESW-YOLO, which has the smallest number of parameters among the comparative models, HLNet demonstrated a 1.7% improvement in mAP_{50} while having

only approximately 40% of the parameters of BESW-YOLO. HLNet's favorable performance balance stemmed from its innovative network architecture, particularly the reconstruction of the neck feature fusion path and the optimized detection head structure. These enhancements boosted detection accuracy while reducing computational scale, thereby making the model more suitable for deployment on computationally embedded systems.

Table 9. Comparison with other SAR ship detection methods on SSDD and HRSID datasets. Bold numbers indicate the best indicators.

Dataset	Method	F1	mAP_{50} (%)	Paras (M)	FLOPs (G)
SSDD	HR-SDNet [45]	0.920	93.7	-	-
	ST-YOLOv8 [46]	0.881	94.3	-	-
	PPA-Net [47]	0.930	95.1	-	-
	BL-Net [48]	0.930	95.2	47.8	41.7
	MHASD [49]	0.950	95.4	5.5	13.7
	AC-YOLO [50]	-	96.7	1.8	**5.4**
	ATSD [51]	0.958	96.8	61.5	7.3
	BESW-YOLO	-	97.3	1.7	-
	ESarDet	**0.960**	97.9	6.2	7.5
	HLNet	0.953	**98.3**	**0.6**	**5.4**
HRSID	DetNet [52]	0.820	83.2	-	-
	CenterNet [53]	0.830	84.6	14.5	61.8
	FCOS	0.850	84.9	31.8	122.9
	ATSD	0.883	87.8	33.6	236.8
	DW-YOLO [54]	0.850	88.8	-	-
	AC-YOLO	-	89.3	1.8	**5.4**
	BESW-YOLO	-	90.0	1.7	-
	CSnNet [55]	-	91.2	12.2	21.7
	Chen' method [56]	-	**91.7**	9.8	93.3
	HLNet	**0.875**	**91.7**	**0.6**	**5.4**

5. Discussion

This section aims to conduct an in-depth analysis of the outstanding performance demonstrated by HLNet, exploring its underlying mechanisms, advantages, and potential limitations.

Experimental results on the SSDD and HRSID datasets demonstrate that HLNet achieves strong detection performance, with mAP_{50} scores of 98.3% and 91.7%, respectively, indicating its competitiveness against mainstream detectors. This performance gain primarily stems from the synergistic design of the proposed modules rather than isolated architectural modifications.

Specifically, the FCM module encodes high-resolution spatial cues from early layers into guidance vectors through channel–spatial interactions, effectively alleviating feature degradation for small targets in deep networks. The MSConv module further enhances robustness to background interference and scale variation by integrating parallel multi-scale convolutions with channel attention, and together they form the MSNet backbone with strong feature retention capability under complex SAR conditions.

At the feature fusion stage, HLPAFPN contributes to both accuracy and efficiency by strengthening propagation of low-level details for small-target detection while streamlining fusion paths and channel dimensions. Finally, HLHead reformulates bounding box regression as distribution modeling by combining grouped convolutions with distribution focal loss, improving robustness to blurred boundaries and cluttered backgrounds with minimal computational overhead.

Extensive ablation studies and heatmap visualizations consistently validate the effectiveness of each component. Collectively, these designs demonstrate that, through

careful architectural coordination, the conventional trade-off between detection accuracy and model complexity can be substantially alleviated, enabling efficient deployment on resource-constrained SAR platforms.

Despite its strong performance in various scenarios, HLNet still has limitations. We observed that for very small vessels that are merged or closely adjacent in SAR images, HLNet occasionally exhibits detection box merging due to the inevitable information loss caused by subsampling (Figure 11). Future work should consider introducing dedicated attention fusion modules tailored for densely packed targets or integrating super-resolution enhancement branches to address these extreme cases further. Additionally, current efforts have focused on algorithmic optimization. Moving forward, comprehensive benchmarking of inference speed and power consumption on specific hardware platforms will be required to bridge the gap between algorithmic innovation and engineering implementation.

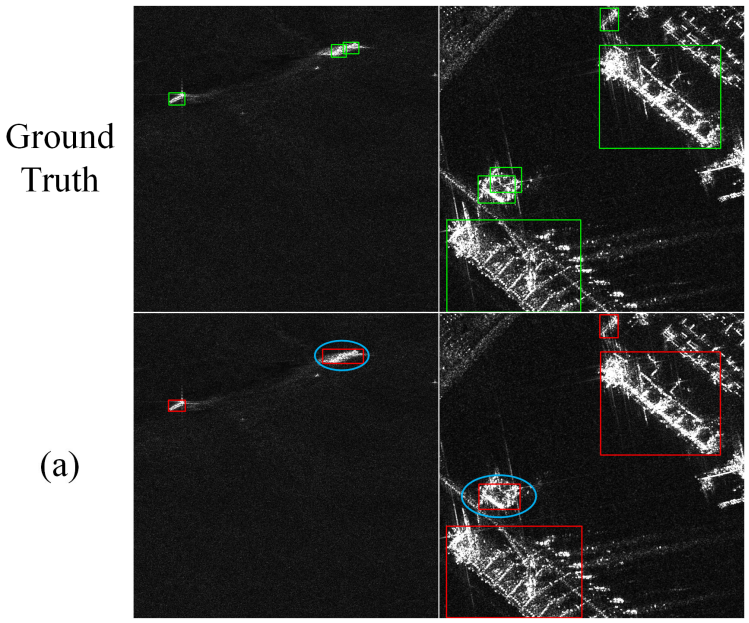


Figure 11. The influence of closely adjacent micro-vessels: (a) shows the output of HLNet. The blue circles represent the merged regions of detection boxes.

6. Conclusions

This paper introduces HLNet, a highly robust and lightweight network for ship detection in complex SAR environments. To address challenges such as intricate sea clutter, large variations in ship scales, and industry demands for lightweight solutions, a series of innovative modules were proposed. The FCM module with complementary channel–spatial mapping enhances spatial localization for small targets, while the MSConv improves robustness to interference and efficiently extracts multi-scale ship features, forming the backbone MSNet. To further integrate multi-scale features, the HLPAFPN feature pyramid network was designed, which fuses low-level details for small-target detection and streamlines the network via optimized fusion paths and reduced channel counts. Finally, a lightweight detection head, HLHead, combining grouped convolutions with DFL, strengthens localization accuracy and anti-noise robustness.

Extensive experiments on SSDD and HRSID demonstrate that HLNet achieves mAP_{50} scores of 98.3% and 91.7%, respectively, with only 0.66 M parameters, achieving a remarkable balance between accuracy and model efficiency. On the more challenging CSID subset, HLNet attains an mAP_{50} of 75.9%, outperforming the baseline by 4.3%, highlighting its strong generalization and interference-resistance capabilities. These results indicate that HLNet is well-suited for deployment in resource-constrained scenarios, such as embedded

systems and edge computing platforms. Future research will advance along two primary directions: (1) We will iteratively refine the model architecture of HLNet to explore its potential in interpreting diverse types of SAR imagery and more complex scenarios. (2) We will focus on deploying the model on various hardware-constrained devices to enable real-time and high-precision detection applications for ship targets.

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

Funding: This research was gratefully supported by the Open Fund of National Engineering Research Center of Geographic Information System, China University of Geosciences, Wuhan 430074, China (Grant No. 2023KFJJ12), the Hubei Key Research and Development Program (Grant No. 2023DJC154), and the Open Fund of Hunan Provincial Key Laboratory of Geo-Information Engineering in Surveying, Mapping and Remote Sensing, Hunan University of Science and Technology (No. E2240).

Data Availability Statement: Our code will be made publicly available at https://github.com/aimu2/HLNet upon acceptance of the manuscript (accessed on 9 February 2026).

Conflicts of Interest: Author Jie Gong was employed by the Wuhan Huaxin Lianchuang Technology Engineering Co., Ltd. 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. Qu, J.; Liu, R.W.; Guo, Y.; Lu, Y.; Su, J.; Li, P. Improving Maritime Traffic Surveillance in Inland Waterways Using the Robust Fusion of AIS and Visual Data. *Ocean Eng.* **2023**, *275*, 114198. [CrossRef]
2. Kurekin, A.; Loveday, B.; Clements, O.; Quartly, G.; Miller, P.; Wiafe, G.; Agyekum, K.A. Use of Sentinel-l and Sentinel-2 for Monitoring Illegal Fishing Off Ghana. In *IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018*; IEEE: New York, NY, USA, 2018; pp. 6875–6878.
3. Wang, T.; Lyu, F.; Chen, C.; Jing, Q.; Han, L. A Behavior Recognition Model for Fishing Vessels Using Limited AIS Trajectory Dataset. *Intell. Mar. Technol. Syst.* **2025**, *3*, 36. [CrossRef]
4. Zhang, C.; Zhang, X.; Gao, G.; Lang, H.; Liu, G.; Cao, C.; Song, Y.; Guan, Y.; Dai, Y. Development and Application of Ship Detection and Classification Datasets: A Review. *IEEE Geosci. Remote Sens. Mag.* **2024**, *12*, 12–45. [CrossRef]
5. 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]
6. Gupta, H.; Verma, O.P.; Sharma, T.K.; Varshney, H.; Agarwal, S.; Pak, W. Ship Detection Using Ensemble Deep Learning Techniques from Synthetic Aperture Radar Imagery. *Sci. Rep.* **2024**, *14*, 29397. [CrossRef]
7. Meester, M.J.; Baslamisli, A.S. SAR Image Edge Detection: Review and Benchmark Experiments. *Int. J. Remote Sens.* **2022**, *43*, 5372–5438. [CrossRef]
8. Zha, C.; Min, W.; Han, Q.; Li, W.; Xiong, X.; Wang, Q.; Zhu, M. SAR Ship Localization Method with Denoising and Feature Refinement. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106444. [CrossRef]
9. Zhou, K.; Zhang, M.; Wang, H.; Tan, J. Ship Detection in SAR Images Based on Multi-Scale Feature Extraction and Adaptive Feature Fusion. *Remote Sens.* **2022**, *14*, 755. [CrossRef]
10. Wang, X.; Xu, W.; Huang, P.; Tan, W. MSSD-Net: Multi-Scale SAR Ship Detection Network. *Remote Sens.* **2024**, *16*, 2233. [CrossRef]
11. 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]
12. Robey, F.C.; Fuhrmann, D.R.; Kelly, E.J.; Nitzberg, R. A CFAR Adaptive Matched Filter Detector. *IEEE Trans. Aerosp. Electron. Syst.* **1992**, *28*, 208–216. [CrossRef]
13. Gandhi, P.P.; Kassam, S.A. Optimality of the Cell Averaging CFAR Detector. *IEEE Trans. Inf. Theory* **1994**, *40*, 1226–1228. [CrossRef]
14. Hansen, V.G.; Sawyers, J.H. Detectability Loss Due to "Greatest Of" Selection in a Cell-Averaging CFAR. *IEEE Trans. Aerosp. Electron. Syst.* **1980**, *AES–16*, 115–118. [CrossRef]
15. Smith, M.E.; Varshney, P.K. VI-CFAR: A Novel CFAR Algorithm Based on Data Variability. In *1997 IEEE National Radar Conference, Syracuse, NY, USA, 13–15 May 1997*; IEEE: New York, NY, USA, 1997; pp. 263–268.

16. Gao, G.; Shi, G. CFAR Ship Detection in Nonhomogeneous Sea Clutter Using Polarimetric SAR Data Based on the Notch Filter. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4811–4824. [CrossRef]
17. Wu, S.; Wang, W.; Deng, J.; Quan, S.; Ruan, F.; Guo, P.; Fan, H. Nearshore Ship Detection in PolSAR Images by Integrating Superpixel-Level GP-PNF and Refined Polarimetric Decomposition. *Remote Sens.* **2024**, *16*, 1095. [CrossRef]
18. Li, J.; Chen, J.; Cheng, P.; Yu, Z.; Yu, L.; Chi, C. A Survey on Deep-Learning-Based Real-Time SAR Ship Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 3218–3247. [CrossRef]
19. Zhang, X.; Zhang, S.; Sun, Z.; Liu, C.; Sun, Y.; Ji, K.; Kuang, G. Semantic Scattering Graph Structure Alignment for Cross-Sensor SAR Image Target Detection. *Chin. J. Aeronaut.* **2026**, 104066. [CrossRef]
20. Sun, Z.; Leng, X.; Zhang, X.; Wu, X.; Xiong, B.; Ji, K.; Kuang, G. KFIA-Net: A Knowledge Fusion and Imbalance-Aware Network for Multi-Category SAR Ship Detection. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *146*, 105127. [CrossRef]
21. 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. [CrossRef]
22. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2016; Volume 9905, pp. 21–37.
23. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*; IEEE: New York, NY, USA, 2016; pp. 779–788.
24. Zhao, T.; Wang, Y.; Li, Z.; Gao, Y.; Chen, C.; Feng, H.; Zhao, Z. Ship Detection with Deep Learning in Optical Remote-Sensing Images: A Survey of Challenges and Advances. *Remote Sens.* **2024**, *16*, 1145. [CrossRef]
25. Mou, F.; Fan, Z.; Ge, Y.; Wang, L.; Li, X. An Efficient Ship Detection Method Based on YOLO and Ship Wakes Using High-Resolution Optical Jilin1 Satellite Imagery. *Sensors* **2024**, *24*, 6708. [CrossRef]
26. Yasir, M.; Liu, S.; Pirasteh, S.; Xu, M.; Sheng, H.; Wan, J.; De Figueiredo, F.A.P.; Aguilar, F.J.; Li, J. YOLOShipTracker: Tracking Ships in SAR Images Using Lightweight YOLOv8. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *134*, 104137. [CrossRef]
27. Xiao, Y.; Xu, T.; Xin, Y.; Li, J. FBRT-YOLO: Faster and Better for Real-Time Aerial Image Detection. *Proc. AAAI Conf. Artif. Intell.* **2025**, *39*, 8673–8681. [CrossRef]
28. Li, X.; Lv, C.; Wang, W.; Li, G.; Yang, L.; Yang, J. Generalized Focal Loss: Towards Efficient Representation Learning for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 3139–3153. [CrossRef] [PubMed]
29. Yang, X.; Xue, S.; Li, L.; Li, S.; Fang, Y.; Zhang, X.; Huang, X. PAMFPN: Position-Aware Multi-Kernel Feature Pyramid Network with Adaptive Sparse Attention for Robust Object Detection in Remote Sensing Imagery. *Remote Sens.* **2025**, *17*, 2213. [CrossRef]
30. Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. SAR Ship Detection Dataset (SSDD): Official Release and Comprehensive Data Analysis. *Remote Sens.* **2021**, *13*, 3690. [CrossRef]
31. 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]
32. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
33. Lin, Z.; Ji, K.; Leng, X.; Kuang, G. Squeeze and Excitation Rank Faster R-CNN for Ship Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 751–755. [CrossRef]
34. Guo, H.; Yang, X.; Wang, N.; Song, B.; Gao, X. A Rotational Libra R-CNN Method for Ship Detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 5772–5781. [CrossRef]
35. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.
36. Zhou, L.; Wan, Z.; Zhao, S.; Han, H.; Liu, Y. BFEA: A SAR Ship Detection Model Based on Attention Mechanism and Multiscale Feature Fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 11163–11177. [CrossRef]
37. Xia, R.; Chen, J.; Huang, Z.; Wan, H.; Wu, B.; Sun, L.; Yao, B.; Xiang, H.; Xing, M. CRTransSar: A Visual Transformer Based on Contextual Joint Representation Learning for SAR Ship Detection. *Remote Sens.* **2022**, *14*, 1488. [CrossRef]
38. Yu, C.; Shin, Y. SAR Ship Detection Based on Improved YOLOv5 and BiFPN. *ICT Express* **2024**, *10*, 28–33. [CrossRef]
39. Zhang, J.; Deng, F.; Wang, Y.; Gong, J.; Liu, Z.; Liu, W.; Zeng, Y.; Chen, Z. DEPDet: A Cross-Spatial Multiscale Lightweight Network for Ship Detection of SAR Images in Complex Scenes. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 18182–18198. [CrossRef]
40. Cao, Q.; Chen, H.; Wang, S.; Wang, Y.; Fu, H.; Chen, Z.; Liang, F. LH-YOLO: A Lightweight and High-Precision SAR Ship Detection Model Based on the Improved YOLOv8n. *Remote Sens.* **2024**, *16*, 4340. [CrossRef]
41. Li, X.; Chen, P.; Yang, J.; An, W.; Zheng, G.; Luo, D.; Lu, A.; Wang, Z. TKP-Net: A Three Keypoint Detection Network for Ships Using SAR Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 364–376. [CrossRef]

42. Zhang, Y.; Chen, C.; Hu, R.; Yu, Y. ESarDet: An Efficient SAR Ship Detection Method Based on Context Information and Large Effective Receptive Field. *Remote Sens.* **2023**, *15*, 3018. [CrossRef]
43. Tang, X.; Cao, K.; Xia, Y.; Cui, E.; Zhao, W.; Chen, Q. BESW-YOLO: A Lightweight SAR Image Detection Model Based on YOLOv8n for Complex Scenarios. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 16081–16094. [CrossRef]
44. Jiang, P.-T.; Zhang, C.-B.; Hou, Q.; Cheng, M.-M.; Wei, Y. LayerCAM: Exploring Hierarchical Class Activation Maps for Localization. *IEEE Trans. Image Process.* **2021**, *30*, 5875–5888. [CrossRef]
45. Wei, S.; Su, H.; Ming, J.; Wang, C.; Yan, M.; Kumar, D.; Shi, J.; Zhang, X. Precise and Robust Ship Detection for High-Resolution SAR Imagery Based on HR-SDNet. *Remote Sens.* **2020**, *12*, 167. [CrossRef]
46. Gao, F.; Tian, Y.; Wu, Y.; Zhang, Y. ST-YOLOv8: Small-Target Ship Detection in SAR Images Targeting Specific Marine Environments. *Appl. Sci.* **2025**, *15*, 6666. [CrossRef]
47. Tang, G.; Zhao, H.; Claramunt, C.; Zhu, W.; Wang, S.; Wang, Y.; Ding, Y. PPA-Net: Pyramid Pooling Attention Network for Multi-Scale Ship Detection in SAR Images. *Remote Sens.* **2023**, *15*, 2855. [CrossRef]
48. Wei, X.; Zhang, X.; Wang, S.; Cheng, C.; Huang, Y.; Yang, K.; Li, Y. BLNet: A Fast Deep Learning Framework for Low-Light Image Enhancement with Noise Removal and Color Restoration. *arXiv* **2021**, arXiv:2106.15953. [CrossRef]
49. Yu, N.; Ren, H.; Deng, T.; Fan, X. A Lightweight Radar Ship Detection Framework with Hybrid Attentions. *Remote Sens.* **2023**, *15*, 2743. [CrossRef]
50. He, R.; Han, D.; Shen, X.; Han, B.; Wu, Z.; Huang, X. AC-YOLO: A Lightweight Ship Detection Model for SAR Images Based on YOLO11. *PLoS ONE* **2025**, *20*, e0327362. [CrossRef]
51. Yao, C.; Xie, P.; Zhang, L.; Fang, Y. ATSD: Anchor-Free Two-Stage Ship Detection Based on Feature Enhancement in SAR Images. *Remote Sens.* **2022**, *14*, 6058. [CrossRef]
52. Li, Z.; Peng, C.; Yu, G.; Zhang, X.; Deng, Y.; Sun, J. DetNet: A Backbone Network for Object Detection. *arXiv* **2018**, arXiv:1804.06215. [CrossRef]
53. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. CenterNet: Keypoint Triplets for Object Detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.
54. Chen, Y.; Zheng, W.; Zhao, Y.; Song, T.H.; Shin, H. DW-YOLO: An Efficient Object Detector for Drones and Self-Driving Vehicles. *Arab. J. Sci. Eng.* **2023**, *48*, 1427–1436. [CrossRef]
55. Chen, C.; Zeng, W.; Zhang, X.; Zhou, Y. CS^{n}Net: A Remote Sensing Detection Network Breaking the Second-Order Limitation of Transformers With Recursive Convolutions. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [CrossRef]
56. Cao, S.; Zhao, C.; Dong, J.; Fu, X. Ship Detection in Synthetic Aperture Radar Images under Complex Geographical Environments, Based on Deep Learning and Morphological Networks. *Sensors* **2024**, *24*, 4290. [CrossRef]

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.