

Article

Not peer-reviewed version

Attention-Guided Multi-Scale Feature Aggregation for Robust SAR Ship Detection in Complex Port Environments

[Xiaokang Qin](#) , [Hongsheng Xu](#) ^{*} , [Zhaodi Wang](#) , Yike Liu , Shuqiang Yang

Posted Date: 9 February 2026

doi: 10.20944/preprints202602.0660.v1

Keywords: SAR ship detection; YOLOv8; multi-scale feature fusion; lightweight attention; port monitoring



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

Attention-Guided Multi-Scale Feature Aggregation for Robust SAR Ship Detection in Complex Port Environments

Xiaokang Qin ¹, Hongsheng Xu ^{1,*}, Zhaodi Wang ², Yike Liu ² and Shuqiang Yang ²

¹ College of Electronic Commerce, Luoyang Normal University, No.6 Jiqing Road, Yibin District, Luoyang, 471934, Henan, China

² Henan Province Engineering Research Center of Robot Advanced Control and Intelligent Information Processing, Luoyang Normal University, No.6 Jiqing Road, Yibin District, Luoyang, 471934, Henan, China

* Correspondence: xhsls@lynu.edu.cn

Abstract

To address the challenges of synthetic aperture radar (SAR) ship detection in complex port environments, including strong background clutter, large-scale variations, and frequent missed detections of small vessels, this paper proposes a lightweight attention-guided multi-scale fusion detector, termed LAMFDet, based on YOLOv8. The proposed framework enhances the baseline architecture from three complementary perspectives: multi-scale feature enhancement, lightweight adaptive feature extraction, and efficient feature reconstruction. Specifically, a multi-scale feature enhancement module (SPPFSENetV2) integrating spatial pyramid pooling with channel attention is designed to strengthen scale-aware contextual representations. A lightweight adaptive extraction (LAE) module is introduced to improve small-target perception while maintaining computational efficiency. Furthermore, an efficient upsampling convolution block (EUCB) is incorporated to preserve structural details and enhance feature fusion quality across feature levels. Extensive experiments on two public SAR ship datasets, HRSID and SSDD, demonstrate the effectiveness and generalization capability of LAMFDet. On HRSID, LAMFDet achieves 94.79% Precision and 71.53% mAP@0.5:0.95, surpassing YOLOv8n by 2.81% and 14.92%, respectively. On SSDD, it further attains 99.65% Precision and 82.05% mAP@0.5:0.95, indicating strong robustness under diverse SAR imaging conditions. These results confirm that LAMFDet effectively improves detection completeness and localization accuracy for small and densely distributed vessels while maintaining favorable efficiency, highlighting its practical potential for real-time port monitoring and maritime surveillance.

Keywords: SAR ship detection; YOLOv8; multi-scale feature fusion; lightweight attention; port monitoring

1. Introduction

Synthetic Aperture Radar (SAR) [1–3] is an active microwave remote sensing technology that enables stable all-weather and day-and-night imaging by acquiring target backscattering information. Compared with optical imaging systems dependent on solar illumination, SAR exhibits superior robustness to cloud cover, fog, and adverse meteorological conditions, and has therefore been widely applied in ocean monitoring, disaster assessment, environmental surveillance, and maritime traffic supervision[4–6]. In maritime applications, these characteristics make SAR one of the most reliable sensing modalities for ship monitoring and maritime situational awareness.

With the rapid growth of global maritime transportation and the digitalization of port logistics systems, automated and fine-grained vessel monitoring in port areas has become increasingly critical.

As key hubs connecting maritime routes and terrestrial logistics networks, ports play a decisive role in ensuring the safety and efficiency of global supply chains[7,8]. Consequently, SAR-based automatic ship detection has emerged as a fundamental technology for port traffic regulation, berth management, illegal activity monitoring, and emergency response, attracting sustained attention from both academia and industry[9,10].

However, SAR ship detection in port and coastal environments remains significantly more challenging than in open seas. Dense strong-scattering man-made structures in ports often exhibit scattering characteristics similar to ships, leading to frequent false alarms [11,12]. Meanwhile, densely distributed vessels with large scale variations—especially small and distant ships occupying only a few pixels—are prone to feature attenuation in deep networks [13,14]. Furthermore, coherent speckle noise inherent to SAR imaging degrades edge and texture information, limiting the effectiveness of detection models originally designed for optical imagery [15,16]. These factors make accurate and robust detection of multi-scale ship targets in complex port scenes a persistent and challenging problem in SAR image interpretation.

1.1. Related Work

Early SAR ship detection studies mainly relied on statistical modeling approaches, among which constant false alarm rate (CFAR) detectors and their variants have been most widely investigated. These methods distinguish ship targets from background clutter by modeling sea clutter distributions and adaptively determining detection thresholds. Wang [17] proposed a CFAR-based ship detection method using range-compressed data for spaceborne SAR systems, improving detection sensitivity while maintaining computational efficiency. Li [18] introduced an adaptive CFAR framework that integrates intensity and texture features with an attention-based contrast mechanism, enhancing target-background discrimination in moderately complex scenes. To improve robustness under heterogeneous clutter conditions, Gao and Liu[19] compared multi-parameter statistical clutter models and demonstrated their advantages over conventional single-parameter models, although their adaptability in complex nearshore environments remains limited. Yuan [20] proposed an adaptive ship detection strategy transferable from optical to SAR imagery by dynamically adjusting detection windows, while Sun [21] employed multilevel super pixel segmentation and fuzzy fusion to reduce false alarms near coastlines. Despite their low computational complexity and ease of implementation, CFAR-based and statistical methods heavily depend on prior assumptions about background clutter distributions. In complex port and nearshore environments characterized by dense man-made structures and strong speckle noise, these assumptions are frequently violated, resulting in degraded detection accuracy and limited robustness, particularly for small and densely distributed ship targets[22,23].

With the rapid development of deep learning, convolutional neural network (CNN)-based object detection methods [24] have become the dominant paradigm for SAR ship detection. Existing deep learning-based detectors can generally be divided into two-stage and one-stage frameworks. Two-stage detectors, such as Faster R-CNN [25,26], generate region proposals followed by refined classification and regression, achieving high localization accuracy, particularly for small targets. One-stage detectors, represented by YOLO [27] and SSD [28], unify localization and classification into an end-to-end framework, significantly improving inference speed while maintaining competitive detection accuracy. To better adapt generic detection frameworks to SAR ship detection tasks, numerous task-oriented improvements have been introduced. Attention mechanisms, such as SENet [29,30], CBAM[31] and non-local attention, have been employed to enhance feature discrimination and suppress background clutter. Orientation-aware detection strategies, including rotated bounding boxes and polar-coordinate representations, have been proposed to address the arbitrary orientations of ships in SAR images [32,33]. Furthermore, multi-scale feature fusion structures, such as FPN [34] and BiFPN [35], are widely adopted to mitigate scale variation issues, while lightweight network designs—such as Ghost convolution [36], depthwise separable convolution [37], and knowledge

distillation [38]—have been explored to reduce computational complexity and facilitate deployment on resource-constrained platforms.

Despite these advances, several limitations persist in complex port scenarios. Small ship targets still suffer from feature attenuation in deep layers, the introduction of sophisticated modules often increases inference latency, and dedicated architectural optimization tailored to SAR imaging mechanisms, particularly speckle noise characteristics, remains insufficiently explored.

1.2. Motivation

Although substantial progress has been achieved in SAR ship detection through both traditional statistical approaches and deep learning-based methods, the related studies reviewed in Section 1.1 indicate that accurate and efficient ship detection in complex port environments remains a challenging problem.

From the perspective of traditional CFAR-based and statistical methods, their effectiveness largely relies on accurate prior modeling of background clutter distributions. However, as demonstrated in [17,18,21], port and nearshore scenes are characterized by strong scattering from dense man-made structures and highly heterogeneous clutter, which significantly deviates from ideal statistical assumptions. As a result, these methods tend to suffer from elevated false alarm rates and poor robustness, especially when detecting small and densely distributed vessels.

With the adoption of deep learning-based detectors, detection accuracy has been notably improved. Nevertheless, existing CNN-based SAR ship detection methods still exhibit several inherent limitations when applied to complex port scenarios. First, although multi-scale feature fusion structures such as FPN and BiFPN have been introduced [34,35], small ship targets with weak scattering responses are still prone to feature attenuation in deep layers, leading to frequent missed detections. Second, attention-based enhancement strategies [29,30,39] and orientation-aware designs [32,33] often introduce additional computational overhead, making it difficult to achieve an effective balance between detection accuracy and real-time inference. Third, most existing methods directly adopt network architectures designed for optical imagery and lack dedicated optimization for SAR-specific imaging characteristics, particularly the influence of coherent speckle noise on feature representation and reconstruction.

YOLOv8 [40], as a state-of-the-art single-stage detector, provides a favorable balance between detection accuracy and inference efficiency, making it a promising baseline for practical SAR ship detection applications. However, its original architecture is not explicitly tailored to the characteristics of SAR imagery or the complex scattering conditions in port environments. In particular, limitations remain in multi-scale feature representation, small-target preservation, and feature reconstruction quality during up-sampling.

Motivated by these observations, this work aims to enhance the YOLOv8 framework through targeted, SAR-aware architectural improvements, focusing on (1) strengthening multi-scale ship feature representation under complex backgrounds, (2) enhancing small-target perception while maintaining network lightness, and (3) improving feature reconstruction quality during feature fusion. By addressing these challenges in a unified and lightweight manner, the proposed method seeks to achieve robust, accurate, and efficient SAR ship detection in complex port environments.

1.3. Our Work

To address the challenges of SAR ship detection in complex port environments, including severe background clutter, large-scale variations, and frequent missed detections of small vessels, we propose a lightweight attention-guided multi-scale fusion detector, termed LAMFDet, built upon the YOLOv8 framework. Different from conventional feature enhancement strategies, LAMFDet is carefully designed to jointly improve multi-scale representation, small-target perception, and feature reconstruction efficiency. The main contributions of this work are summarized as follows:

- We propose a multi-scale feature enhancement module (SPPFSENetV2) that integrates spatial pyramid pooling with channel attention. This module explicitly strengthens scale-aware

contextual modeling and adaptively emphasizes ship-related features, enabling more robust discrimination of vessels under complex port backgrounds.

- We design a lightweight adaptive extraction (LAE) module to improve sensitivity to small and weak ship targets while maintaining computational efficiency. By adaptively refining shallow features with minimal overhead, LAE effectively balances detection accuracy and real-time performance.

- We introduce an efficient upsampling convolution block (EUCB) to enhance feature reconstruction during the decoding stage. By combining depthwise separable convolution with channel shuffle operations, EUCB preserves structural details and boundary information, significantly reducing missed detections in dense and small-object scenarios.

- Extensive experiments conducted on two public SAR ship datasets, HRSID and SSDD, demonstrate that LAMFDet consistently outperforms mainstream detection models in terms of detection accuracy, robustness, and generalization capability, while maintaining favorable inference efficiency. These results validate the effectiveness of the proposed framework and highlight its practical potential for real-time port monitoring and maritime surveillance applications.

2. Method

To address the challenges commonly encountered in SAR-based port monitoring scenarios—including strong sea clutter interference, highly imbalanced ship scale distributions, and frequent missed detections of small vessels—we propose a lightweight attention-guided multi-scale fusion detector, termed LAMFDet, built upon the single-stage YOLOv8 framework. By introducing targeted structural designs at critical stages of the detection pipeline, including feature extraction, feature fusion, and feature reconstruction, the proposed method effectively enhances detection accuracy and robustness for multi-scale SAR ship targets while maintaining a lightweight architecture.

YOLOv8 consists of three main components: a backbone for feature extraction, a neck for multi-scale feature fusion, and a detection head, as illustrated in Figure 1. Owing to its end-to-end design, YOLOv8 achieves a favorable balance between accuracy and inference efficiency in optical image detection tasks. However, when directly applied to SAR port scenes, the baseline architecture exhibits notable limitations. Specifically, strong sea clutter and complex man-made structures interfere with high-level semantic discrimination; small-scale vessels tend to be suppressed in deeper feature representations; and conventional feature fusion and upsampling strategies are insufficient for effectively recovering fine-grained structural details. These issues collectively lead to degraded localization accuracy and a high false-negative rate for small and densely distributed ships.

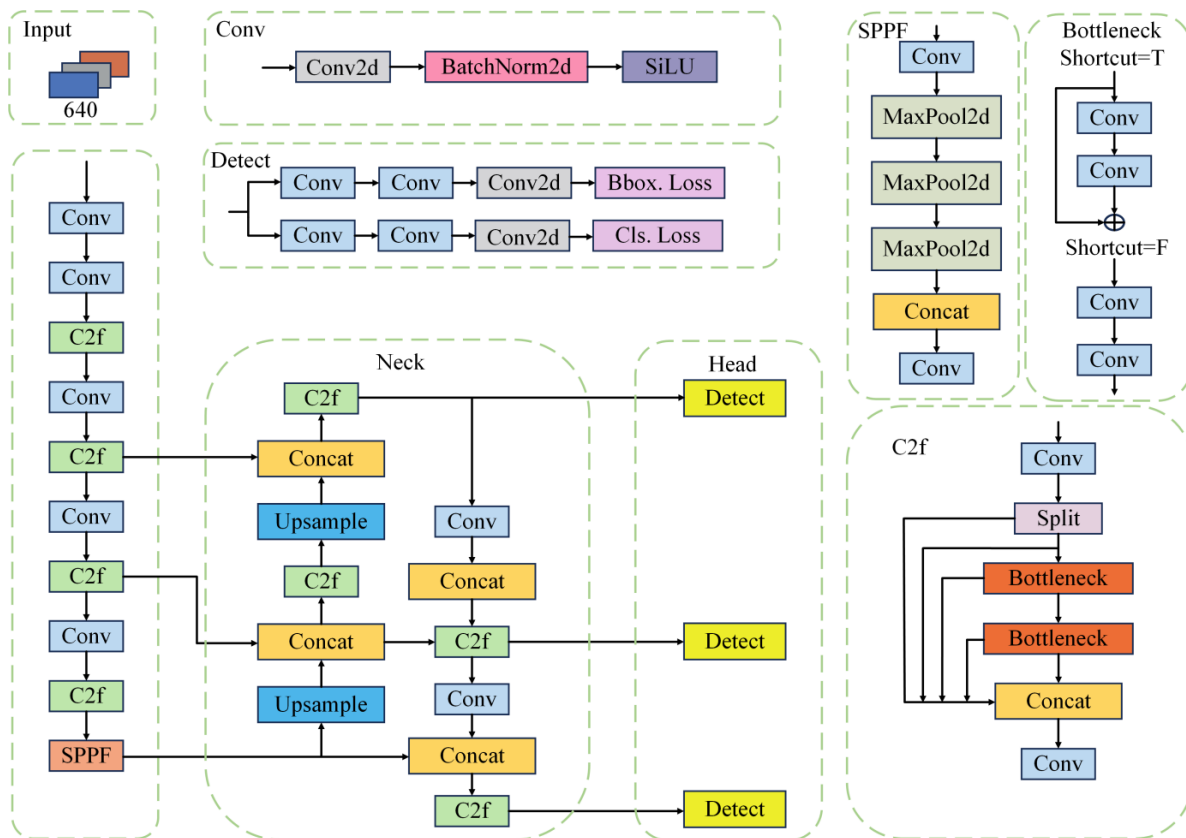


Figure 1. YOLOv8 network structure.

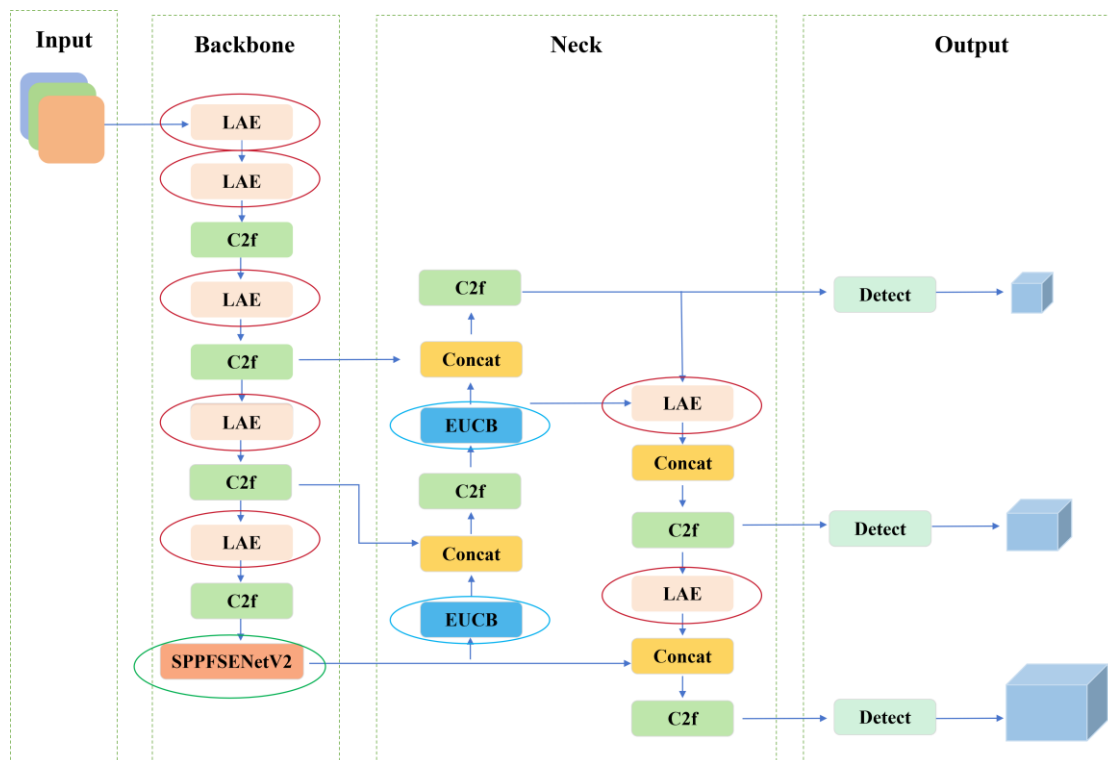


Figure 2. Proposed LAMFDet network architecture.

To overcome these limitations, the overall architecture of LAMFDet is illustrated in Figure 2. Rather than simply increasing network depth or channel width, this work focuses on the intrinsic characteristics of SAR ship detection in complex port environments and introduces coordinated optimizations at three critical stages of information flow within the network.

First, during high-level semantic feature extraction, a multi-scale feature enhancement module, termed SPPFSENetV2, is introduced to strengthen scale-aware contextual modeling by integrating spatial pyramid pooling with channel attention. This design enables the network to better capture vessels of varying sizes while suppressing background clutter.

Second, at shallow and intermediate feature levels, we design a lightweight adaptive extraction (LAE) module to enhance local feature representation for small ships while maintaining computational efficiency. By adaptively refining low-level features, LAE improves small-target perception without introducing significant overhead.

Third, during feature fusion and reconstruction, an efficient upsampling convolution block (EUCB) is incorporated to preserve structural details and improve feature recovery quality. By combining depthwise separable convolution with channel shuffle operations, EUCB enhances boundary representation and mitigates information loss caused by conventional upsampling, thereby reducing missed detections in dense and small-object scenarios.

Through these coordinated multi-stage optimizations, LAMFDet preserves the efficiency advantages of single-stage detectors while substantially improving detection completeness and localization accuracy for multi-scale SAR ship targets. Consequently, the proposed framework demonstrates strong adaptability and practical deployment potential in complex coastal and port monitoring applications.

2.1. Lightweight Adaptive Feature Extraction Module

In SAR port and near-shore imaging scenarios, small vessels constitute a large proportion of detected targets and are frequently distributed in dense clusters. Owing to their limited spatial extent, such targets typically occupy only a few pixels, imposing stringent requirements on the feature representation capability of shallow and intermediate network layers. Although increasing network depth or channel width can enhance feature discrimination, these strategies inevitably introduce substantial parameter overhead and computational burden, thereby limiting applicability in real-time and resource-constrained deployment environments.

To address this issue, we design a Lightweight Adaptive Enhancement (LAE) feature extraction module that aims to improve small-target representation while preserving computational efficiency. Rather than relying on heavy convolutional operations, LAE adopts depthwise separable convolution to decouple spatial and channel-wise feature learning, significantly reducing parameter complexity while maintaining effective receptive field coverage.

As illustrated in Figure 3, the LAE module further incorporates an adaptive feature fusion mechanism to dynamically emphasize informative responses associated with small vessels. By selectively enhancing discriminative local features and suppressing background interference, LAE strengthens sensitivity to weak scattering ship targets in cluttered port scenes.

Through this lightweight yet adaptive design, LAE effectively improves small-scale vessel perception at early feature stages without increasing model depth, enabling a better balance between detection accuracy and inference efficiency. This module plays a critical role in reducing missed detections of small ships while maintaining the real-time performance characteristics of the overall LAMFDet framework.

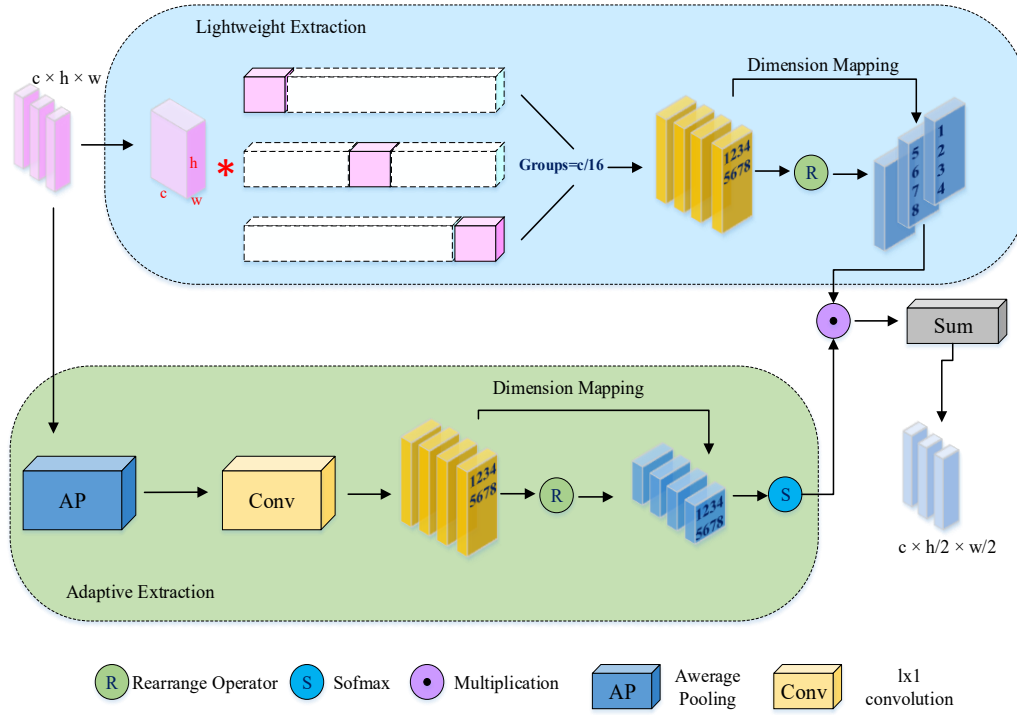


Figure 3. LAE Module Structural Diagram.

Let the input feature map be denoted as:

$$X \in \mathbb{R}^{H \times W \times C} \quad (1)$$

where H , W , and C represent the height, width, and number of channels of the feature map, respectively.

Depthwise Separable Convolution. To efficiently decouple spatial and channel-wise feature learning, the LAE module first applies depthwise separable convolution, which consists of a depthwise convolution followed by a pointwise convolution. The computation can be expressed as:

$$Y = Conv_{pw}(Conv_{dw}(X)) \quad (2)$$

where $Conv_{dw}(\cdot)$ denotes depthwise convolution that independently models local spatial information for each channel, and $Conv_{pw}(\cdot)$ denotes pointwise (1×1) convolution used to fuse information across channels.

Compared with standard convolution, this structure significantly reduces computational cost and parameter redundancy while effectively preserving fine-grained spatial details, which is particularly beneficial for small-scale ship targets occupying limited pixels in SAR images.

Adaptive Channel Weighting Mechanism. After local spatial feature modeling, the LAE module introduces a lightweight adaptive channel weighting mechanism to further enhance the discriminative capability of ship-related features. Specifically, global average pooling (GAP) is first applied to the intermediate feature map Y to extract global channel-wise statistics:

$$z_c = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W Y_{i,j,c}, \quad c = 1, 2, \dots, C \quad (3)$$

where $z \in \mathbb{R}^C$ is the channel descriptor vector.

Subsequently, the channel descriptor is mapped using a 1×1 convolution (or equivalently, a fully connected layer) parameterized by learnable weights W , followed by a Softmax function to generate normalized channel attention weights:

$$w = Sotfmax(W_z) \quad (4)$$

where $w \in \mathbb{R}^C$ satisfies $\sum_{c=1}^C w_c = 1$.

Finally, the original feature map Y is recalibrated through channel-wise weighting to obtain the output feature map F :

$$F_c = w_c \cdot Y_c, \quad c = 1, 2, \dots, C \quad (5)$$

Through this adaptive weighting mechanism, the LAE module dynamically amplifies channel responses that are highly correlated with ship targets while suppressing background clutter and noise-induced features. As a result, small-scale ship targets become more distinguishable in the feature space, leading to improved detection performance without introducing significant computational overhead.

2.2. Improved Multi-Scale Feature Enhancement Module

In SAR port imaging scenarios, ship targets exhibit significant scale variation. Small vessels usually occupy only a limited number of pixels, whereas large ships present more distinctive geometric structures and scattering patterns. Conventional single-scale feature extraction strategies struggle to model such wide scale distributions simultaneously, particularly in deep networks where repeated downsampling operations tend to suppress or even eliminate fine-grained representations of small targets. These limitations severely degrade detection performance in complex port environments.

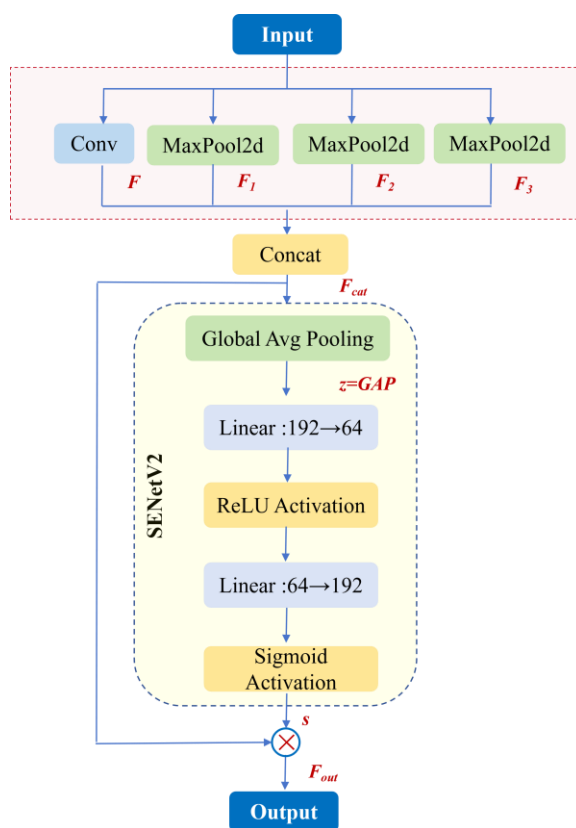


Figure 4. SPPFSENetV2 module structure diagram.

To alleviate this problem, we design an improved multi-scale feature enhancement module, termed SPPFSENetV2, and integrate it into the backbone of YOLOv8. The overall structure of SPPFSENetV2 is illustrated in Figure 4. By jointly exploiting multi-scale spatial pyramid pooling and channel-wise attention, the proposed module enhances scale-aware contextual modeling while suppressing background interference.

Let the input feature map be denoted as

$$F \in \mathbb{R}^{H \times W \times C} \quad (6)$$

where C , H , and W represent the number of channels, height, and width of the feature map, respectively.

Multi-Scale Spatial Pyramid Pooling. To capture contextual information under different receptive fields, the input feature map F is first processed by multiple max-pooling branches with different kernel sizes. The resulting multi-scale feature representations are expressed as:

$$F_i = \text{MaxPool}_{k_i}(F), \quad i = 1, 2, \dots, N \quad (7)$$

where k_i denotes the pooling kernel size corresponding to the i -th scale, and $N = 3$ is the number of pooling branches. Each branch extracts spatial features at a distinct receptive field, enabling effective modeling of ship targets with varying spatial extents.

Subsequently, the multi-scale pooled features are concatenated with the original input feature along the channel dimension to form a fused feature representation:

$$F_{cat} = \text{Concat}(F, F_1, F_2, F_3, \dots, F_N) \quad (8)$$

Channel Attention via Squeeze-and-Excitation. To further enhance discriminative channel responses, a Squeeze-and-Excitation (SE) attention mechanism is applied to the fused feature map F_{cat} . First, global average pooling (GAP) is employed to aggregate spatial information into a channel-wise descriptor:

$$z = \text{GAP}(F_{cat}), \quad z \in \mathbb{R}^{C'} \quad (9)$$

where C' denotes the number of channels in the fused feature map.

The channel descriptor is then passed through two fully connected layers (implemented as 1×1 convolutions) with non-linear activation to generate channel-wise attention weights:

$$s = \sigma(W_2 \delta(W_1 z)) \quad (10)$$

where $\delta(\cdot)$ and $\sigma(\cdot)$ denote the ReLU and Sigmoid activation functions, respectively, and W_1 and W_2 are learnable parameters.

Finally, the fused feature map is recalibrated through channel-wise multiplication to obtain the output feature representation:

$$F_{out} = s \odot F_{cat} \quad (11)$$

where \odot denotes element-wise channel-wise multiplication.

By integrating multi-scale spatial pyramid pooling with SE-based channel attention, SPPFSENetV2 enables adaptive emphasis on scale-relevant ship features while suppressing redundant scattering responses from complex port backgrounds. This design significantly enhances high-level semantic robustness and provides strong scale-aware representations for subsequent detection stages, especially benefiting small and densely distributed vessel targets.

2.3. Improved Efficient Up-Sampling Block

In object detection networks, up-sampling plays a critical role during feature fusion by restoring spatial resolution and recovering target structural information. However, conventional up-sampling strategies, such as interpolation-based methods or transposed convolution, often introduce feature blurring and amplify redundant noise in SAR imagery. These effects are particularly detrimental to small-scale ship targets, whose weak scattering responses and limited spatial extent make them highly sensitive to detail degradation during feature reconstruction.

To address these limitations, we propose an Efficient Up-Sampling Convolution Block (EUCB), which aims to improve structural feature recovery and spatial detail preservation during multi-level feature fusion while maintaining low computational overhead. By enhancing cross-scale feature

consistency, EUCB effectively reduces missed detections of small vessels in complex port environments.

The overall architecture of EUCB is illustrated in Figure 5. The core design integrates depthwise separable convolution with a channel shuffle mechanism, enabling efficient spatial refinement and inter-channel information exchange after up-sampling. This design enhances boundary continuity and fine-grained structural representations without introducing excessive parameters, making it suitable for real-time SAR ship detection.

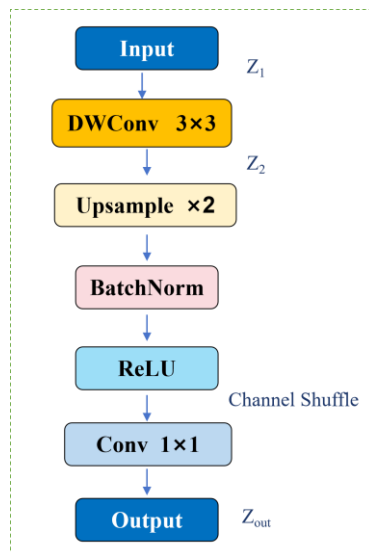


Figure 5. The Efficient Up-sampling Module (EUCB) structure diagram.

Let the input feature map be denoted as $Z \in \mathbb{R}^{H \times W \times C}$, where C , H , and W represent the number of channels, height, and width of the feature map, respectively.

First, a depth wise separable convolution is applied to remap the input features, aiming to reduce computational complexity while enhancing local spatial modeling capability. This process can be formulated as:

$$Z_1 = Conv_{pw}(Conv_{dw}(Z)) \quad (12)$$

where $Conv_{dw}(\cdot)$ denotes the depthwise convolution that independently models spatial information within each channel, and $Conv_{pw}(\cdot)$ denotes the pointwise (1×1) convolution used to facilitate inter-channel feature interaction. This decomposition significantly reduces parameter count and FLOPs while preserving fine-grained spatial details, which is particularly beneficial for SAR ship targets.

Subsequently, spatial resolution is restored through an up-sampling operation:

$$Z_2 = Up(Z_1) \quad (13)$$

where $Up(\cdot)$ denotes nearest-neighbor or bilinear interpolation. Unlike conventional pipelines that directly up-sample high-level semantic features, EUCB performs feature remapping prior to up-sampling, allowing semantic and structural information to be better preserved during resolution recovery.

To further enhance feature expressiveness, a channel shuffle operation is introduced:

$$Z_{out} = Shuffle(Z_2) \quad (14)$$

where $Shuffle(\cdot)$ redistributes feature channels across different groups, alleviating channel isolation caused by depthwise or grouped convolutions and enabling more effective cross-channel interaction in subsequent fusion stages.

By jointly exploiting depthwise separable convolution, structure-aware up-sampling, and channel shuffle, EUCB enhances boundary continuity and local structural consistency of ship targets. This design substantially improves feature reconstruction quality and reduces missed detections of small and densely distributed vessels, contributing to more robust SAR ship detection in complex port environments.

2.4. Loss Function and Optimization Strategy

To ensure fair comparison with baseline methods and isolate the effectiveness of the proposed architectural improvements, LAMFDet adopts the standard loss formulation of YOLOv8 without modification. A multi-task joint optimization strategy is employed to simultaneously supervise bounding box regression, target classification, and localization quality estimation. The overall training objective is defined as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{box} + \lambda_2 \mathcal{L}_{cls} + \lambda_3 \mathcal{L}_{dfl} \quad (15)$$

where \mathcal{L}_{box} , \mathcal{L}_{cls} , and \mathcal{L}_{dfl} denote the bounding box regression loss, classification loss, and Distribution Focal Loss (DFL), respectively. The coefficients λ_1 , λ_2 , and λ_3 balance the contributions of each task.

For bounding box regression, considering the large-scale variation and blurred edges of SAR ship targets, an IoU-based loss is adopted to geometrically constrain the overlap between predicted boxes B_p and ground-truth boxes B_{gt} .

$$\mathcal{L}_{box} = 1 - IoU(B_p, B_{gt}) \quad (16)$$

which improves robustness against large scale variations and ambiguous boundaries commonly observed in SAR ship targets.

For target classification, Binary Cross-Entropy (BCE) loss is adopted:

$$\mathcal{L}_{cls} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (17)$$

where p_i represents the predicted probability of the i -th sample being a ship, $y_i \in \{0,1\}$ is the corresponding ground-truth label, and N is the number of samples. The BCE loss demonstrates numerical stability in class-imbalanced scenarios, making it suitable for SAR ship detection tasks.

To further enhance localization precision, Distribution Focal Loss (DFL) is introduced to model bounding box offsets as discrete probability distributions, enabling more accurate regression of continuous coordinates:

$$\mathcal{L}_{dfl} = -\sum_{k=1}^K y_k \log(p_k) \quad (18)$$

where p_k denotes the predicted probability of the k -th discrete interval, y is the true bounding box value, and K is the number of discrete intervals.

During training, the AdamW optimizer is employed for parameter updating, combined with a cosine annealing learning rate schedule to promote stable convergence and improved generalization.

Notably, all training hyperparameters and loss configurations follow the default YOLOv8 settings, ensuring that performance gains arise solely from the proposed multi-scale feature aggregation and attention-guided architectural enhancements.

3. Experimental Results

To comprehensively evaluate the effectiveness, robustness, and practical applicability of the proposed method for SAR ship detection in complex port environments, extensive experiments were conducted on publicly available SAR ship datasets. The experimental evaluation consists of

quantitative comparisons with state-of-the-art detectors, ablation studies to analyze the contribution of individual modules, and qualitative visual inspections to assess detection performance under challenging scenarios. All experiments were performed under a unified hardware and software configuration to ensure fairness, reproducibility, and reliable performance comparison.

3.1. Experimental Setup

All experiments were conducted under a unified hardware and software environment to ensure fair comparison and reproducibility. The proposed method was implemented using Python 3.9.20 based on the PyTorch 2.1.2 framework, with GPU acceleration enabled via CUDA 11.8. Experiments were performed on a workstation equipped with an NVIDIA GeForce RTX 4060Ti GPU (16 GB memory) running the Windows operating system. Detailed platform configurations are summarized in Table 1.

Table 1. Experimental environment configuration.

Environment Configuration	Parameter
Operating system	Windows
GPU	NVIDIA GeForce RTX 4060Ti 16G
Development environment	PyCharm 2023.3.7
Language	Python 3.9.20
frame	PyTorch 2.1.2
Operating platform	CUDA 11.8

During training, all SAR images were uniformly resized to 640×640 pixels, with a batch size of 16. The AdamW optimizer was employed with an initial learning rate of 0.001, momentum of 0.937, and weight decay of 0.0005. A cosine annealing learning rate scheduler was adopted, together with a warm-up strategy of 3 epochs, to promote stable convergence. Model weights were randomly initialized without using external pre-trained parameters, ensuring that performance gains originated solely from the proposed architectural improvements.

To enhance model generalization in complex SAR environments, diverse data augmentation strategies were applied, including random rotation, translation, scaling, horizontal flipping, Mosaic augmentation, Copy-Paste, RandAugment, and random erasing (probability 0.4). Mosaic augmentation was disabled during the final 10 epochs to improve localization stability. Automatic mixed precision (AMP) training was enabled to accelerate convergence and reduce memory consumption. All experiments were conducted with a fixed random seed (42) and deterministic settings to guarantee reproducibility. Detailed training hyperparameters are provided in **Table 2**.

During inference, the confidence threshold and non-maximum suppression (NMS) threshold were set to 0.25 and 0.45, respectively. The lightweight YOLOv8n model was selected as the baseline detector to comprehensively evaluate the effectiveness and efficiency of the proposed LAMFDet framework under identical training conditions.

Table 2. Experimental parameter settings.

Name	Value
Image size	640640
Batch size	16
Pretraining weight	0
Optimizer	AdamW
Learning rate	0.001
Weight decay	0.0005
Learning Rate Scheduling	Cosine
Momentum	0.937
label_smoothing	0.0

auto_augment	randaugment
erasing	0.4
close_mosaic	10

3.3. Dataset Description

To comprehensively evaluate the effectiveness and generalization capability of the proposed LAMFDet, experiments were conducted on two publicly available SAR ship detection benchmarks: the High-Resolution SAR Ship Detection Dataset (HRSID) [41] and the SAR Ship Detection Dataset (SSDD) [41]. The parameter settings and data partitioning information for both datasets are summarized in **Table 3**.

Table 3. Parameter configuration for the dataset

Dataset	Images	Ships	Classes	Image Size	Split Ratio
HRSID	5,604	16,951	1	640 × 640	7:2:1
SSDD	1,160	2,587	1	~500 × 500	8:2

HRSID is a widely used benchmark for high-resolution SAR ship detection, containing 5,604 SAR images with a spatial size of 640 × 640 pixels and resolutions ranging from 0.5 m to 3 m. The dataset was collected from multiple SAR platforms, including TerraSAR-X, Sentinel-1, and TanDEM-X, covering diverse imaging modes and observation conditions. It provides annotations for 16,951 ship targets distributed across open sea, coastal, and complex port environments. According to the proportion of bounding box area relative to the image size, targets are categorized into small, medium, and large vessels. Small ships dominate the dataset, accounting for approximately 54.5% of all instances, while medium and large vessels represent 43.5% and 2%, respectively. This highly imbalanced scale distribution, together with dense ship layouts and strong background clutter in port scenes, poses significant challenges for multi-scale feature representation and small-target detection.

Figure 6 illustrates representative examples of ships at different scales, including large, medium, and small vessels, highlighting the significant intra-class variation and small-object dominance in HRSID. Following common practice, the dataset was randomly split into training, validation, and test sets with a ratio of 7:2:1, enabling comprehensive evaluation of detection accuracy, robustness, and generalization performance, particularly for small-scale ships in complex port scenarios.

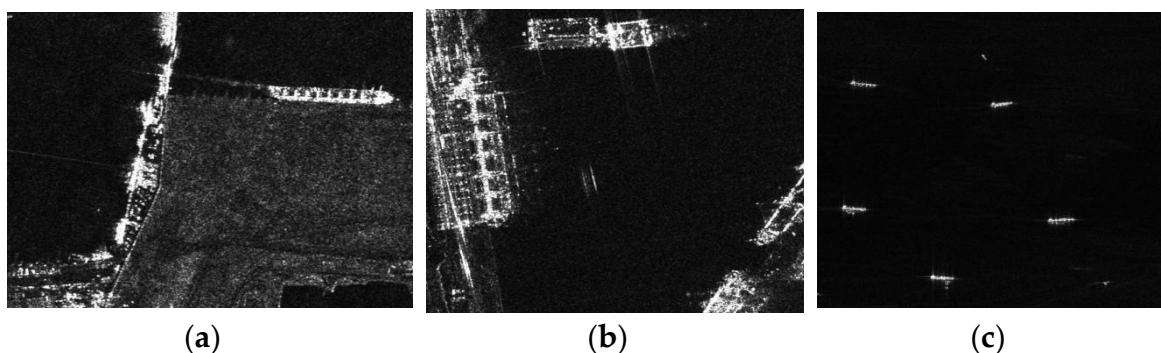


Figure 6. Examples of ship segments: (a) large ship; (b) medium-sized ship; (c) small ship.

SSDD provides complementary evaluation conditions with relatively simpler offshore backgrounds and sparser ship distributions. It consists of 1,160 SAR images with approximately 2,587 annotated ship instances, where all images have an approximate spatial resolution of 500 × 500 pixels and only a single target category (ship) is considered. The dataset was randomly divided into training

and test subsets using an 8:2 ratio. During training on SSDD, models were optimized for 400 epochs to ensure full convergence.

By jointly evaluating on HRSID and SSDD, which exhibit distinct scene complexities and target distributions, the proposed method is systematically assessed in both challenging port environments and relatively unconstrained maritime scenarios, enabling a comprehensive analysis of detection accuracy, robustness, and generalization performance.

3.3. Evaluation Metrics

To comprehensively evaluate the detection performance and practical applicability of the proposed method in complex port SAR scenarios, multiple quantitative metrics are adopted from different perspectives, including detection accuracy, robustness, and computational efficiency. Specifically, mAP@0.5, mAP@0.5:0.95, Precision, Recall, FPS, Params, and FLOPs are used as evaluation indicators, which conform to widely accepted standards in object detection and SAR ship detection research.

Precision (P) and Recall (R) are first employed to characterize the basic detection behavior. Precision measures the reliability of predicted ship targets, while Recall reflects the capability of detecting all existing ships. They are defined as.

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN} \quad (19)$$

where TP , FP , and FN denote the numbers of true positives, false positives, and false negatives, respectively. In port SAR scenes, high Precision indicates effective suppression of false alarms caused by strong-scattering man-made structures, whereas high Recall is critical for reducing missed detections of small and densely distributed vessels.

The primary accuracy metric is mean Average Precision (mAP). For a given IoU threshold τ , the Average Precision (AP) is computed as the area under the Precision–Recall curve:

$$AP_{\tau} = \int_0^1 P(r) dr \quad (20)$$

mAP@0.5 corresponds to the mean AP with an IoU threshold of 0.5, reflecting the overall detection capability. To further evaluate localization robustness, mAP@0.5:0.95 averages AP values over multiple IoU thresholds from 0.5 to 0.95 with a step of 0.05, providing a more stringent assessment of bounding box accuracy, which is especially important in cluttered port environments.

In addition to detection performance, computational efficiency is evaluated to assess deployment feasibility. FPS (Frames Per Second) measures inference speed under a unified hardware setting. Params (M) denote the number of trainable parameters, and FLOPs (G) quantify the computational complexity of the model. Lower Params and FLOPs indicate better suitability for real-time or resource-constrained maritime surveillance platforms.

3.4. Comparative Experiment Results and Analysis

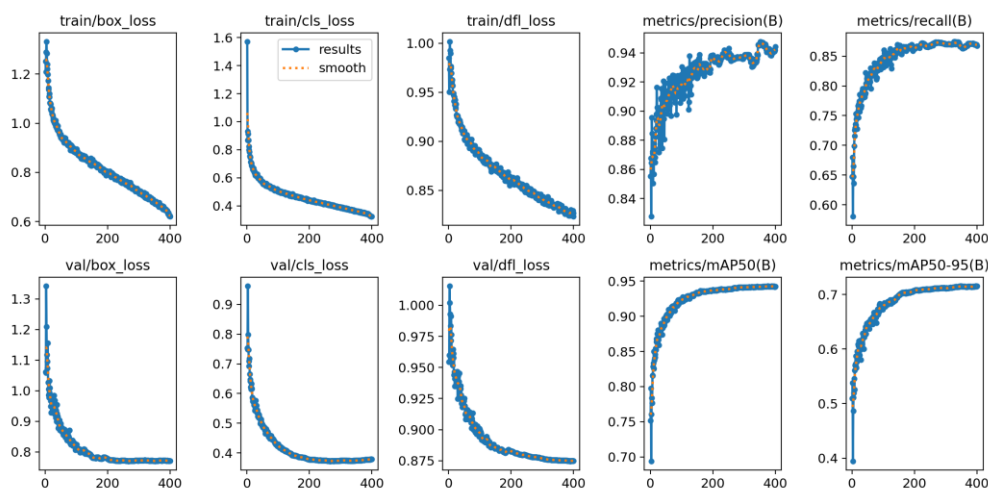
To comprehensively evaluate the detection accuracy, convergence behavior, and generalization capability of the proposed LAMFDet, extensive comparative experiments were conducted on both the HRSID and SSDD datasets. The evaluation protocol consists of three components: (1) training convergence and classification reliability analysis, (2) quantitative comparison with representative detection models and cross-dataset validation, and (3) qualitative visualization of detection results. Specifically, training dynamics and confusion matrices are first analyzed to assess optimization stability and classification performance. Subsequently, quantitative comparisons are presented to demonstrate accuracy–efficiency trade-offs and generalization ability across datasets. Finally, visual detection results on both datasets further illustrate the practical advantages of the proposed method in complex SAR scenes.

3.4.1. Training Convergence and Classification Analysis

Figure 7 presents the training curves of LAMFDet on both HRSID and SSDD datasets, illustrating the evolution of Precision, Recall, mAP@0.5, and validation classification loss. The consistent convergence behavior across datasets indicates strong optimization stability and dataset-independent learning capability. As observed, the proposed model exhibits rapid convergence on both datasets, with all evaluation metrics increasing sharply during the early training stage and gradually stabilizing after approximately 30–40 epochs. This behavior indicates that the introduced multi-scale feature enhancement and lightweight adaptive extraction mechanisms effectively facilitate discriminative feature learning for SAR ship targets.

On both datasets, Precision and Recall curves show steady upward trends with only minor oscillations, demonstrating stable optimization without evident overfitting. Meanwhile, the mAP@0.5 curves consistently converge to high values, confirming the effectiveness of the proposed architecture in modeling ships with large scale variations under complex SAR scattering conditions. Furthermore, the smooth and continuous decrease of validation classification loss verifies the numerical stability and strong generalization capability of LAMFDet across different SAR imaging environments.

To further assess classification reliability, the normalized confusion matrices on HRSID and SSDD are presented in **Figure 8**. As shown in **Figure 8(a)**, on the more challenging HRSID dataset, LAMFDet achieves a true positive rate of approximately 93% for ship targets, with only about 7% misclassified as background. This result demonstrates strong discriminative capability in cluttered port environments characterized by dense ship distributions and complex coastal structures. In contrast, near-perfect classification performance is observed on SSDD (**Figure 8(b)**), where ship targets are almost entirely correctly identified, reflecting the robustness of the proposed model under relatively simpler sea-surface backgrounds.



(a)

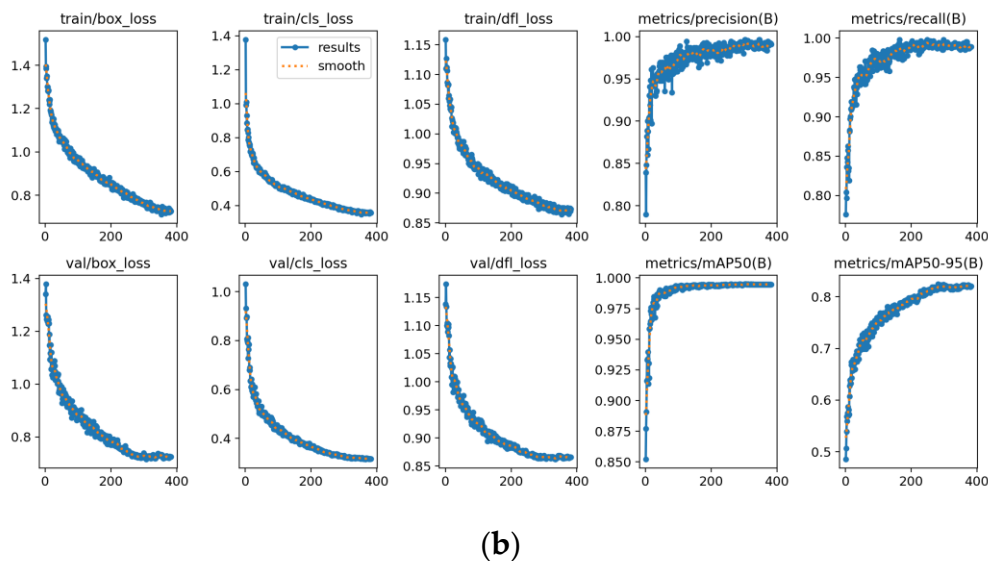


Figure 7. Training curves of the proposed LAMFDet on different datasets: (a) HRSID; (b) SSDD.

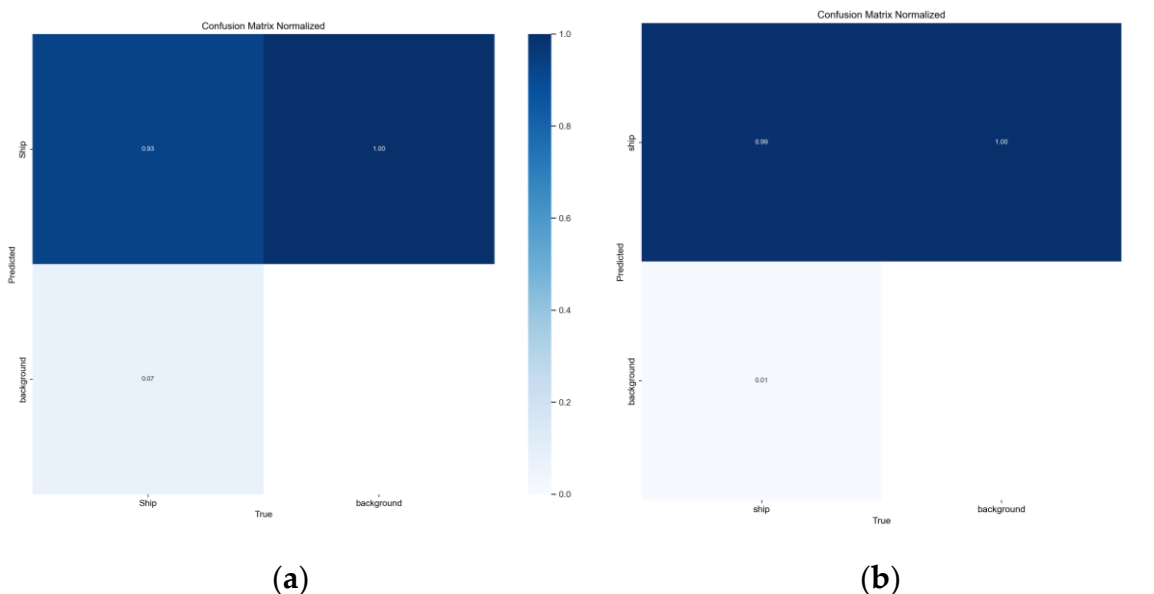


Figure 8. Normalized confusion matrices of LAMFDet on different datasets: (a) HRSID; (b) SSDD.

3.4.2. Quantitative Comparison and Cross-Dataset Evaluation

To evaluate detection accuracy, computational efficiency, and cross-dataset generalization, LAMFDet was compared with representative two-stage and one-stage detectors on the HRSID and SSDD datasets. All YOLO-based models were retrained under identical settings, whereas results of other methods were adopted from their original reports to ensure fair comparison.

(1) Performance Comparison on HRSID

Table 4 summarizes the quantitative results on HRSID in terms of Precision, Recall, mAP@0.5, mAP@0.5:0.95, inference speed, parameter size, and FLOPs. Traditional two-stage detectors exhibit limited robustness in complex SAR port environments. Faster R-CNN achieves only 86.82% mAP@0.5 and 54.15% mAP@0.5:0.95, indicating insufficient localization accuracy under dense ship distributions and strong background clutter. RetinaNet provides marginal improvement but remains constrained at higher IoU thresholds, revealing limited adaptability to scale variation and SAR-specific scattering characteristics. Among lightweight one-stage detectors, YOLOv5s and YOLOv7-tiny improve mAP@0.5 to 89.61% and 90.17%, respectively, yet their Recall remains below 86%, suggesting persistent missed detections of small or weakly scattering vessels. YOLOv8n further

increases Precision to 91.98%, but Recall drops to 80.64%, reflecting the inherent limitations of ultra-lightweight architectures in modeling complex harbor scenes.

With increased network capacity, YOLOv8s significantly improves Recall and mAP@0.5:0.95 to 87.27% and 60.14%, respectively. YOLOv11s further elevates mAP@0.5:0.95 to 66.86%, demonstrating the benefit of deeper feature representations. Attention-enhanced variants (YOLOv11s + CBAM and YOLOv11s + LAE_SPPF) achieve mAP@0.5:0.95 of 68.65% and 69.71%, confirming that feature recalibration and multi-scale contextual aggregation effectively suppress background interference.

In contrast, the proposed LAMFDet achieves the best overall performance, reaching 94.79% Precision, 87.46% Recall, 94.30% mAP@0.5, and 71.53% mAP@0.5:0.95. Compared with the strongest competing method (YOLOv11s + LAE_SPPF), LAMFDet improves mAP@0.5:0.95 by 1.82% and Precision by 1.51%, while maintaining competitive Recall. These results demonstrate that the coordinated integration of multi-scale feature aggregation and lightweight adaptive enhancement substantially improves localization accuracy and small-target sensitivity.

Notably, LAMFDet preserves real-time inference capability (82 FPS on RTX 4060Ti) with moderate computational overhead (14.8 M parameters and 37.5 GFLOPs), achieving a favorable balance between accuracy and efficiency.

Table 4. Comparative experimental results of different algorithms.

Models	Precision (%)	Recall (%)	mAP_0.5 (%)	mAP_0.5:0.95 (%)	FPS (RTX 4060)	Params (M)	FLOPs (G)
Faster R-CNN [42]	85.68	82.32	86.82	54.15	18	41.3	172.6
RetinaNet [43]	86.27	83.13	87.56	55.03	22	38.5	142.3
YOLOv5s [39]	90.82	82.94	89.61	57.46	82	7.2	16.5
YOLOv7-tiny [44]	89.47	85.63	90.17	58.29	118	6.1	13.2
YOLOv8n [45]	91.98	80.64	90.34	56.61	142	3.2	8.7
YOLOv8s [46]	92.45	87.27	92.19	60.14	90	11.2	28.6
YOLOv11s [47]	92.93	88.14	92.62	66.86	86	12.6	31.8
YOLOv11s + CBAM [48]	93.44	83.06	92.41	68.65	78	13.3	34.7
YOLOv11s + LAE_SPPF	93.28	83.84	92.88	69.71	74	13.5	36.2
Ours	94.79	87.46	94.30	71.53	82	14.8	37.5

Training Dynamics Analysis. Figure 9 further compares the training dynamics of representative models on HRSID. LAMFDet exhibits faster early-stage performance improvement and converges to consistently higher final values across Precision, Recall, mAP@0.5, and mAP@0.5:0.95. Moreover, its curves show reduced oscillations, indicating improved optimization stability.

LAMFDet surpasses baseline models within approximately the first 20–30 epochs and maintains higher performance plateaus in later stages, whereas YOLOv8n shows noticeable fluctuations and YOLOv11s converges to lower ceilings. These observations suggest that LAMFDet benefits not only from superior final accuracy but also from accelerated convergence and stable training behavior, which are particularly important for SAR ship detection in dense and cluttered port environments.

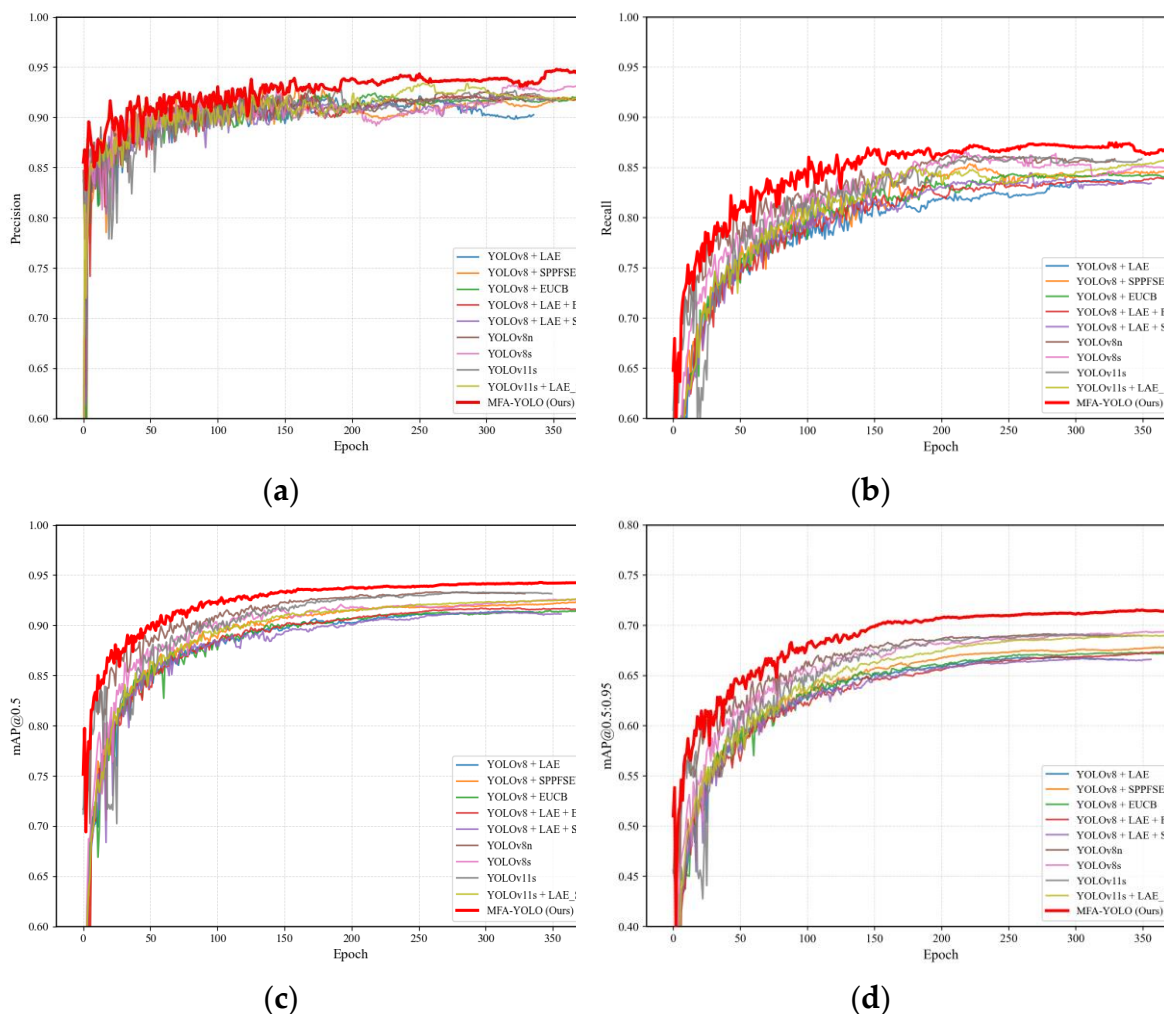


Figure 9. Training performance comparison of representative detection models on the HRSID dataset: (a) Precision curves; (b) Recall curves; (c) mAP@0.5 curves; (d) mAP@0.5:0.95 curves.

(2) Cross-Dataset Evaluation on SSDD

To assess generalization capability, experiments were extended to SSDD, with results summarized in **Table 5**. LAMFDet again achieves superior performance, attaining 99.65% Precision, 99.26% Recall, and 99.50% mAP@0.5, outperforming all comparison methods. Compared with YOLOv8n, LAMFDet simultaneously improves both Precision and Recall, indicating enhanced sensitivity to small vessels under different SAR imaging conditions.

Although SSDD is relatively less complex than HRSID, achieving simultaneously near-saturated Precision and Recall still requires effective multi-scale representation and robust feature discrimination. The observed performance gains therefore reflect the effectiveness of the proposed architectural design rather than favorable data characteristics alone. These results confirm that LAMFDet generalizes well across datasets with different resolutions, acquisition platforms, and scene complexities, without relying on simple model scaling.

Table 5. Quantitative comparison on HRSID and SSDD datasets.

Models	Precision (%)	Recall (%)	mAP_0.5 (%)	Precision (%)	Recall (%)	mAP_0.5 (%)
	HRSID			SSDD		
YOLOv5s [39]	90.82	82.94	89.61	92.36	91.88	92.63
YOLOv7-tiny [44]	89.47	85.63	90.17	95.62	96.45	96.03
YOLOv8n [45]	91.98	80.64	90.34	97.45	96.92	97.85

YOLOv8s [46]	92.45	87.27	92.19	98.13	97.56	98.13
YOLOv11s [47]	92.93	88.14	92.62	99.05	97.98	99.24
YOLOv11s + CBAM [48]	93.44	83.06	92.41	98.28	96.80	98.75.
YOLOv11s + LAE_SPPF	93.28	83.84	92.88	99.26	98.25	99.41
Ours	94.79	87.46	94.30	99.65	99.26	99.50

(3) Precision–Recall Analysis

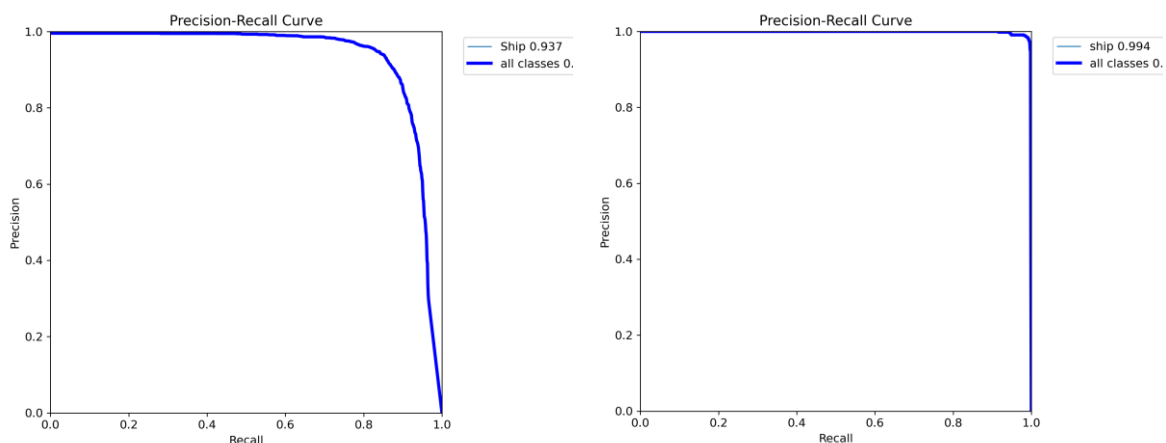
Figure 10 further presents the Precision–Recall (PR) curves of LAMFDet on the HRSID and SSDD datasets, offering an intuitive characterization of detection reliability under varying confidence thresholds. Since both datasets contain a single ship category, the PR curves directly reflect the trade-off between precision and recall for vessel targets.

On HRSID (**Figure 10(a)**), LAMFDet achieves an mAP@0.5 of 0.937, with precision remaining above 0.9 across most of the recall range. As recall approaches saturation, a gradual decline in precision is observed, which mainly results from densely distributed small vessels and strong background interference in complex port scenes. This curve behavior indicates that, although extremely high recall inevitably introduces a limited number of false positives, LAMFDet maintains high confidence predictions over a wide operating region. This observation is consistent with the quantitative results in Table 4, where LAMFDet attains 94.79% Precision, 87.46% Recall, and 94.30% mAP@0.5, confirming its robustness under challenging multi-scale and cluttered SAR environments.

In contrast, on SSDD (**Figure 10(b)**), LAMFDet reaches an mAP@0.5 of 0.994, with the PR curve remaining near the upper-right corner across almost the entire recall interval, indicating near-perfect detection stability. This remarkably high performance can be attributed to several factors. First, SSDD generally contains relatively sparse ship distributions and less complex background structures compared with HRSID, reducing ambiguity between targets and surrounding clutter. Second, ships in SSDD exhibit clearer scattering responses and more consistent scale characteristics, facilitating feature discrimination. More importantly, the proposed multi-scale aggregation and lightweight adaptive extraction mechanisms enable LAMFDet to fully exploit these favorable data properties, yielding simultaneously high precision and recall.

This behavior is quantitatively corroborated by **Table 5**, where LAMFDet achieves 99.65% Precision, 99.26% Recall, and 99.50% mAP@0.5 on SSDD, outperforming all comparison methods. Compared with YOLOv8n, LAMFDet improves both Precision and Recall, demonstrating enhanced sensitivity to small vessels while effectively suppressing false alarms.

Overall, by jointly analyzing Figure 10 together with Tables 4 and 5, it can be concluded that LAMFDet exhibits consistent superiority across both complex (HRSID) and relatively simpler (SSDD) SAR scenarios. The proposed architecture achieves stable high performance through coordinated multi-scale feature modeling and adaptive enhancement, confirming its strong generalization capability and practical applicability for SAR ship detection.



(a) (b)

Figure 10. Precision–Recall (PR) curves of LAMFDet on different datasets: (a) HRSID; (b) SSDD.

3.5. Visualisation of Test Results

To further demonstrate the practical detection capability of the proposed LAMFDet in complex SAR environments, qualitative comparisons were conducted on both HRSID and SSDD datasets. Figures 11–13 present representative detection results under diverse scenarios, including dense port regions, coastal waters, and sparse offshore scenes. For clarity, false positives are marked by yellow circles, while missed detections are indicated by cyan boxes.

Figure 11 shows a typical dense port scene from the HRSID dataset, comparing Ground Truth, YOLOv5, YOLOv8n, and LAMFDet. YOLOv5 exhibits both false alarms and missed detections, including two false positives and six missed vessels. YOLOv8n reduces false alarms but still misses four targets and produces two false detections, mainly involving small or weakly scattering ships. In contrast, LAMFDet detects 49 out of 50 ground-truth vessels, with only a single missed target, demonstrating substantially improved detection completeness and background suppression capability in highly cluttered harbor environments.

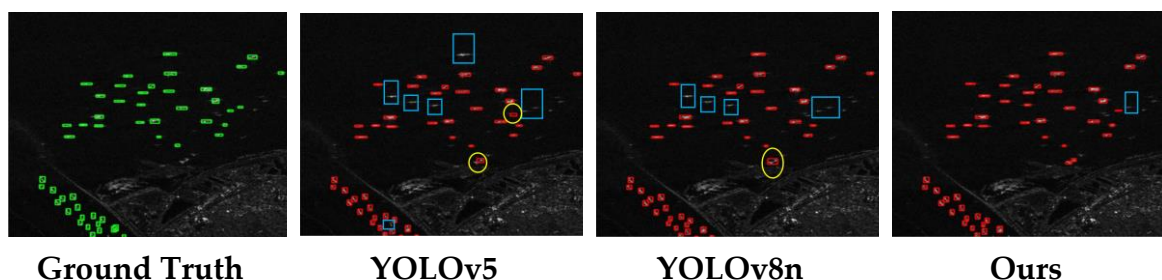
**Figure 11.** Qualitative comparison on a dense port scene from the HRSID dataset.

Figure 12 further illustrates qualitative comparisons on HRSID across four representative scenes. YOLOv8n frequently fails to identify small vessels near coastlines and densely clustered ship regions, and occasionally produces false alarms caused by strong land backscatter or port infrastructures. Benefiting from multi-scale feature aggregation and lightweight adaptive enhancement, LAMFDet consistently recovers these missed targets while effectively suppressing spurious responses, achieving more accurate localization and higher confidence predictions, especially for densely distributed and small-scale vessels.



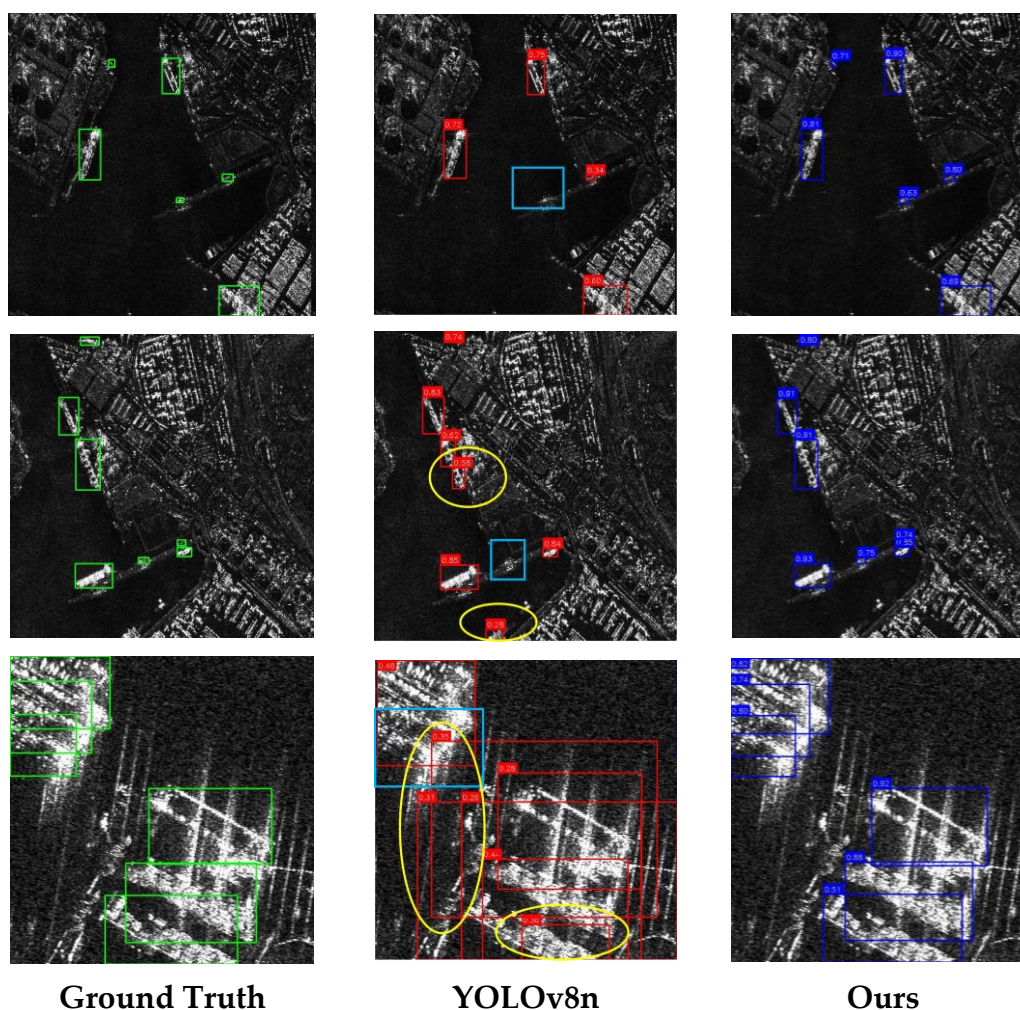
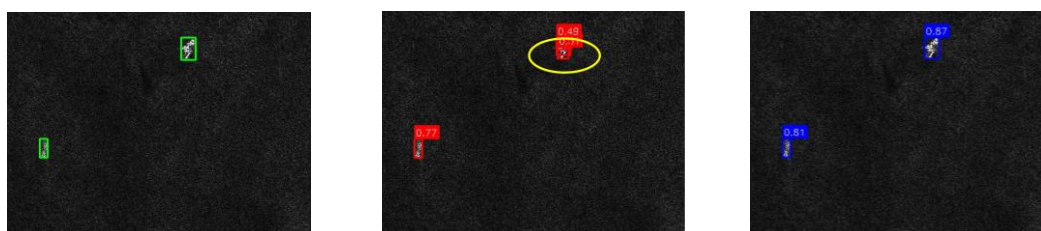


Figure 12. Visualization results on the HRSID dataset under four representative scenes.

Figure 13 presents visualization results on the SSDD dataset under five typical offshore scenarios. Although YOLOv8n already achieves relatively strong performance on SSDD, it still suffers from missed detections and occasional false positives in low-contrast regions. In comparison, LAMFDet produces more complete and cleaner detection results, with improved confidence scores for small and isolated ships. These observations are consistent with the quantitative results in Table 5 and the PR curves in Figure 9, where LAMFDet achieves near-saturated precision and recall on SSDD.

Overall, the qualitative results across both datasets are highly consistent with the quantitative evaluations reported in Tables 4–5 and Figure 9, confirming that LAMFDet effectively enhances detection completeness while suppressing false alarms. The performance gains are particularly evident for small-scale vessels and densely populated port scenes, highlighting the effectiveness of the proposed multi-scale feature modeling and adaptive enhancement strategy.



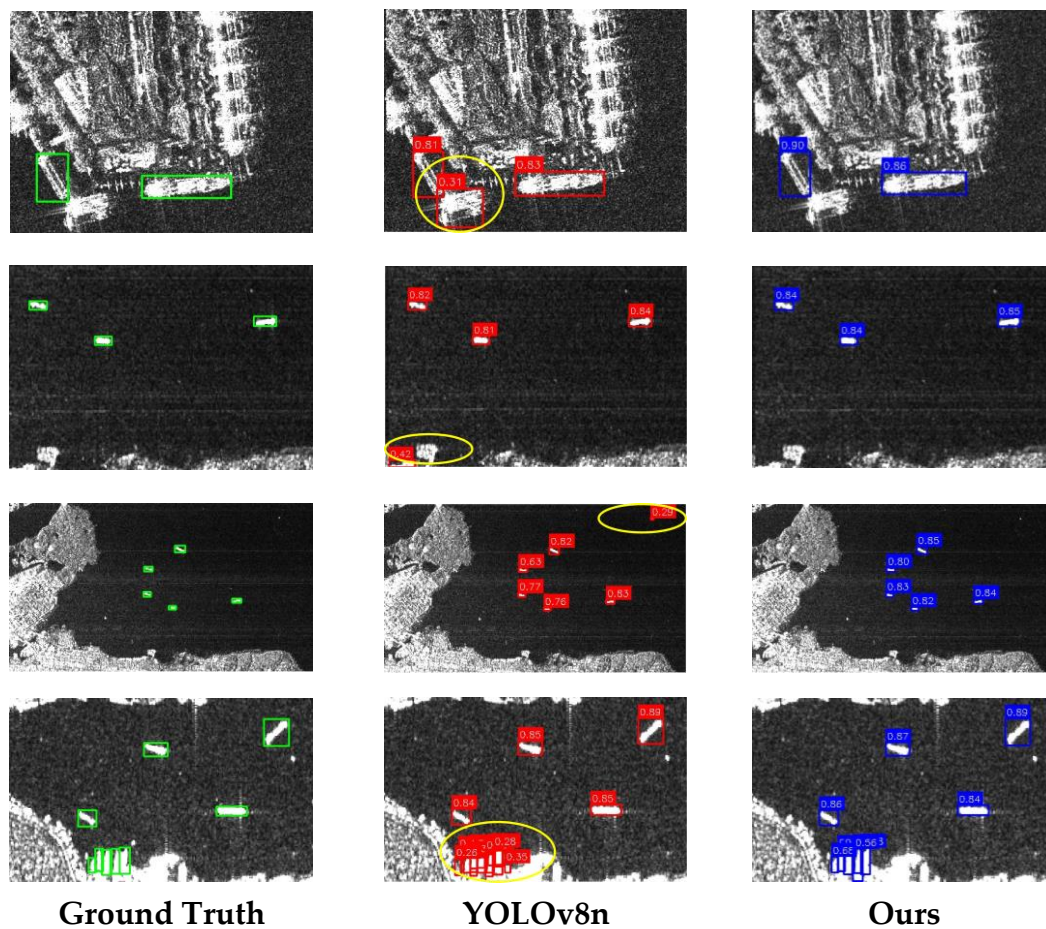


Figure 13. Visualization results on the SSDD dataset under five representative scenes.

3.6. Ablation Experiment

To quantitatively analyze the contribution of each proposed component, ablation experiments were conducted on the HRSID dataset using YOLOv8n as the baseline. The Lightweight Adaptive Enhancement (LAE), SPPFSENetV2, and Enhanced Upsampling Compensation Block (EUCB) were progressively introduced under identical training settings. The results are summarized in Table 6. The baseline YOLOv8n achieves Precision, Recall, mAP@0.5, and mAP@0.5:0.95 of 91.97%, 80.63%, 90.31%, and 65.65%, respectively, indicating noticeable missed detections and limited localization accuracy in dense small-ship scenarios.

Table 6. Ablation experimental results.

YOLOv8n	LAE	SPPFSENetV2	EUCB	Precision (%)	Recall (%)	mAP_0.5 (%)	mAP_0.5:0.95 (%)
√	/	/	/	91.97	80.63	90.31	65.65
√	√	/	/	92.09	82.05	91.06	66.66
√	/	√	/	92.04	78.53	89.05	62.59
√	/	/	√	92.40	82.95	90.94	66.48
√	√	√	/	92.45	82.90	91.25	66.67
√	√	/	√	92.41	81.05	90.18	64.77
√	/	√	√	93.06	84.99	92.59	69.54
√	√	√	√	94.79	87.46	94.30	71.53

Effect of Individual Modules. Introducing each module independently leads to consistent performance improvements: (1) LAE enhances local discriminative feature extraction, increasing

Recall to 82.05% and mAP@0.5 to 91.06%, confirming its effectiveness in strengthening shallow representations for small vessels. (2) EUCB improves feature upsampling quality and semantic compensation, raising Recall to 82.95% and mAP@0.5:0.95 to 66.48%, demonstrating its capability in recovering fine-grained spatial details. (3) SPPFSENetV2 mainly strengthens multi-scale context aggregation; however, when used alone, its improvement is limited, indicating that multi-scale fusion benefits from complementary attention and reconstruction mechanisms.

Synergistic Effects of Module Combinations. When multiple components are combined, clear synergistic effects emerge. In particular, integrating SPPFSENetV2 + EUCB increases Recall and mAP@0.5:0.95 to 84.99% and 69.54%, respectively, confirming the complementarity between multi-scale fusion and enhanced feature reconstruction. Other dual-module configurations also outperform single-module variants, further demonstrating that the proposed components reinforce each other rather than acting independently.

Full Model Performance. With all three modules integrated to form the complete LAMFDet, the model achieves the best overall performance, reaching 94.79% Precision, 87.46% Recall, 94.30% mAP@0.5, and 71.53% mAP@0.5:0.95. Compared with the baseline, mAP@0.5:0.95 improves by 5.88%, indicating a substantial enhancement in high-IoU localization accuracy.

Overall, the ablation results demonstrate that the performance gains of LAMFDet arise from the coordinated collaboration of lightweight adaptive enhancement, multi-scale feature aggregation, and semantic compensation, rather than from any single component alone. This collaborative design effectively strengthens small-target perception and spatial localization, enabling robust detection of multi-scale ships in complex SAR environments.

4. Conclusions

This paper presents LAMFDet, an enhanced YOLOv8-based SAR ship detection framework tailored for complex port environments with dense multi-scale targets and strong background clutter. By integrating SPPFSENetV2 for multi-scale feature enhancement, the Lightweight Adaptive Enhancement (LAE) module for small-target perception, and the Efficient Upsampling Convolution Block (EUCB) for feature reconstruction, LAMFDet effectively improves scale-aware representation, localization accuracy, and detection robustness. Experiments on HRSID and SSDD demonstrate consistent performance gains over representative detectors. LAMFDet achieves 94.30% mAP@0.5 and 71.53% mAP@0.5:0.95 on HRSID, while reaching 99.50% mAP@0.5 on SSDD, confirming strong generalization across datasets with different scene complexities. Meanwhile, real-time inference is maintained with moderate computational overhead. Ablation studies further verify that the improvements originate from the collaborative contribution of all proposed modules rather than any single component.

Future work will focus on extending LAMFDet toward more realistic operational scenarios. First, incorporating temporal SAR sequences or multi-source remote sensing data may further improve robustness against occlusion and weak scattering. Second, adaptive model compression and hardware-aware optimization will be explored to facilitate deployment on satellite-borne and airborne platforms. Finally, improving generalization under extreme weather conditions and highly cluttered coastal environments remains an important direction, aiming to provide reliable technical support for intelligent port management and maritime safety monitoring.

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

Funding: This research was funded in part by the National Natural Science Foundation of China (grant 62301241) , part by the Key Research and Development and Promotion Program (Science and Technology Tackling Project) of Henan Province (Grant No. 242102210085), and in part by the Key Research Program of Higher Education Institutions in Henan Province (Grant No. 25A510017).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Acknowledgments: The authors would like to acknowledge the anonymous reviewers and editors whose thoughtful comments helped to improve this manuscript.

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

References

1. Jiang, D.; Marino, A.; Ionescu, M.; Gvilava, M.; Savaneli, Z.; Loureiro, C.; Spyrakos, E.; Tyler, A.; Stanica, A.J.I.J.o.P.; Sensing, R. Combining optical and SAR satellite data to monitor coastline changes in the Black Sea. *2025*, *226*, 102-115.
2. Jeong, W.; Song, M.-S.; Adhikari, M.D.; Yum, S.-G.J.B. Monitoring the integrity and vulnerability of linear urban infrastructure in a reclaimed coastal city using SAR interferometry. *2025*, *15*, 3865.
3. Li, J.; Xu, C.; Su, H.; Gao, L.; Wang, T.J.R.S. Deep learning for SAR ship detection: Past, present and future. *2022*, *14*, 2712.
4. Haloho, L.S.; Supriyadi, A.A.J.R.S.T.i.D.; Environment. Utilization of satellite technology in communication systems, disaster monitoring, border surveillance, and military intelligence: A literature review. *2024*, *1*, 36-44.
5. Şengül, B.; Yılmaz, F.; Uğurlu, Ö.J.S. Safety–security analysis of maritime surveillance systems in critical marine areas. *2023*, *15*, 16381.
6. Zhang, B.; Wu, Q.; Wu, F.; Huang, J.; Wang, C.J.R.S. A lightweight pyramid transformer for high-resolution SAR image-based building classification in port regions. *2024*, *16*, 3218.
7. Lee, P.T.-W.; Song, Z.-Y.; Lin, C.-W.; Lam, J.S.L.; Chen, J.J.T.R. New framework of port logistics in the post-COVID-19 period with 6th-generation ports (6GP) model. *2025*, *45*, 77-93.
8. Wang, P.; Hu, Q.; Mei, Q.; Wang, S.; Yang, Y.; Guo, D.; Liu, X.; Hu, W.; Chen, J.J.A.E.I. Intelligent port logistics: A spatiotemporal knowledge graph and AI-agent framework for berth allocation. *2025*, *68*, 103633.
9. Zhang, C.; Gao, G.; Zhang, X.; Li, H.-C.; Liu, T.; Zhang, Z.; Li, G.J.I.G.; Magazine, R.S. An overview of polarized synthetic aperture radar ship detection: From traditional methods to artificial intelligence. *2025*.
10. Xue, W.; Ai, J.; Zhu, Y.; Chen, J.; Zhuang, S.J.I.T.o.A.; Systems, E. AIS-FCANet: Long-term AIS data assisted frequency-spatial contextual awareness network for salient ship detection in SAR imagery. *2025*.
11. Zhou, H.; Geng, Z.; Sun, M.; Wu, L.; Yan, H.J.S. Context-Guided SAR Ship Detection with Prototype-Based Model Pretraining and Check–Balance-Based Decision Fusion. *2025*, *25*, 4938.
12. Liu, S.; Li, D.; Song, H.; Fan, C.; Li, K.; Wan, J.; Liu, R.J.I.J.o.P.; Sensing, R. SAR ship detection across different spaceborne platforms with confusion-corrected self-training and region-aware alignment framework. *2025*, *228*, 305-322.
13. Wang, L.; Zhou, Z.; Luo, R.; Zhao, L.; Liu, L.J.I.J.o.S.T.i.A.E.O.; Sensing, R. Small Ship Detection in SAR Images Based on Asymmetric Feature Learning and Shallow Context Embedding. *2025*.
14. Zhu, H.; Li, D.; Wang, H.; Yang, R.; Liang, J.; Liu, S.; Wan, J.J.R.S. A Progressive Saliency-Guided Small Ship Detection Method for Large-Scene SAR Images. *2025*, *17*, 3085.
15. Wang, X.; Wang, Y.; Li, S.J.I.J.o.S.T.i.A.E.O.; Sensing, R. Study on Coherent Speckle Noise Suppression in the SAR Images Based on Regional Division. *2025*.
16. Saha, A.; Singh, H.; Maji, S.K.J.S.P.I.C. Integrated multi-channel approach for speckle noise reduction in SAR imagery using gradient, spatial, and frequency analysis. *2025*, 117406.

17. Wang, C.; Guo, B.; Song, J.; He, F.; Li, C.J.I.T.o.G.; Sensing, R. A novel CFAR-based ship detection method using range-compressed data for spaceborne SAR system. **2024**, *62*, 1-15.
18. Li, N.; Pan, X.; Yang, L.; Huang, Z.; Wu, Z.; Zheng, G.J.S. Adaptive CFAR method for SAR ship detection using intensity and texture feature fusion attention contrast mechanism. **2022**, *22*, 8116.
19. Gao, S.; Liu, H.J.I.J.o.S.T.i.A.E.O.; Sensing, R. Performance comparison of statistical models for characterizing sea clutter and ship CFAR detection in SAR images. **2022**, *15*, 7414-7430.
20. Yuan, Y.; Rao, Z.; Lin, C.; Huang, Y.; Ding, X.J.I.G.; Letters, R.S. Adaptive ship detection from optical to SAR images. **2023**, *20*, 1-5.
21. Sun, Q.; Liu, M.; Chen, S.; Lu, F.; Xing, M.J.I.T.o.G.; Sensing, R. Ship detection in SAR images based on multilevel superpixel segmentation and fuzzy fusion. **2023**, *61*, 1-15.
22. Wang, X.; Li, G.; Plaza, A.; He, Y.J.I.T.o.G.; Sensing, R. Revisiting SLIC: Fast superpixel segmentation of marine SAR images using density features. **2022**, *60*, 1-18.
23. Zhang, L.; Liu, Y.; Zhao, W.; Wang, X.; Li, G.; He, Y.J.I.T.o.G.; Sensing, R. Frequency-adaptive learning for SAR ship detection in clutter scenes. **2023**, *61*, 1-14.
24. Wang, Z.; Yang, S.; Qin, H.; Liu, Y.; Wang, J.J.E. MixCFormer: A CNN-Transformer Hybrid with Mixup Augmentation for Enhanced Finger Vein Attack Detection. **2025**, *14*, 362.
25. Jiang, M.; Gu, L.; Li, X.; Gao, F.; Jiang, T.J.I.T.o.G.; Sensing, R. Ship contour extraction from SAR images based on faster R-CNN and Chan-Vese model. **2023**, *61*, 1-14.
26. Liu, Q.; Ning, M. SAR ship target detection method based on improved faster R-CNN. In Proceedings of the Sixth International Conference on Information Science, Electrical, and Automation Engineering (ISEAE 2024), 2024; p. 1327502.
27. Ren, X.; Bai, Y.; Liu, G.; Zhang, P.J.R.S. YOLO-Lite: An efficient lightweight network for SAR ship detection. **2023**, *15*, 3771.
28. Wen, G.; Cao, P.; Wang, H.; Chen, H.; Liu, X.; Xu, J.; Zaiane, O.J.A.I. MS-SSD: Multi-scale single shot detector for ship detection in remote sensing images. **2023**, *53*, 1586-1604.
29. Deng, H.; Wang, S.; Wang, X.; Zheng, W.; Xu, Y.J.E. YOLO-SEA: An Enhanced Detection Framework for Multi-Scale Maritime Targets in Complex Sea States and Adverse Weather. **2025**, *27*, 667.
30. Cai, S.; Meng, H.; Wu, J.J.J.o.R.-T.I.P. FE-YOLO: YOLO ship detection algorithm based on feature fusion and feature enhancement. **2024**, *21*, 61.
31. Rocha, R.d.L.; Figueiredo, F.A.d.J.R.S. Enhancing YOLO-Based SAR Ship Detection with Attention Mechanisms. **2025**, *17*, 3170.
32. Liu, J.; Liu, L.; Xiao, J.J.I.J.o.S.T.i.A.E.O.; Sensing, R. Ellipse polar encoding for oriented SAR ship detection. **2024**, *17*, 3502-3515.
33. Li, P.; Feng, C.; Feng, W.; Hu, X.J.I.J.o.S.T.i.A.E.O.; Sensing, R. Oriented bounding box representation based on continuous encoding in oriented SAR ship detection. **2025**.
34. Zhang, Z.T.; Zhang, X.; Shao, Z. Deform-FPN: A novel FPN with deformable convolution for multi-scale SAR ship detection. In Proceedings of the IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium, 2023; pp. 5273-5276.
35. Yu, C.; Shin, Y.J.I.E. SAR ship detection based on improved YOLOv5 and BiFPN. **2024**, *10*, 28-33.
36. Chen, J.; Huang, J.; Tan, Y.; Wu, Z.; Luo, R.J.R.S. LGNet: A Lightweight Ghost-Enhanced Network for Efficient SAR Ship Detection. **2025**, *17*, 3800.
37. Shang, R.; He, J.; Wang, J.; Xu, K.; Jiao, L.; Stolkin, R.J.K.-B.S. Dense connection and depthwise separable convolution based CNN for polarimetric SAR image classification. **2020**, *194*, 10542.
38. Liu, Z.; Wang, S.; Li, Y.; Gu, Y.; Yu, Q.J.I.T.o.G.; Sensing, R. Dsrkd: Joint despeckling and super-resolution of sar images via knowledge distillation. **2024**.
39. Wang, Z.; Yang, S.; Qin, H.; Liu, Y.; Ding, J.J.I. CCW-YOLO: A modified YOLOv5s network for pedestrian detection in complex traffic scenes. **2024**, *15*, 762.
40. Sohan, M.; Sai Ram, T.; Rami Reddy, C.V. A review on yolov8 and its advancements. In Proceedings of the International Conference on Data Intelligence and Cognitive Informatics, 2024; pp. 529-545.

41. Cavli, B.; Çinarer, G. YOLO Based Ship Detection in SAR Images: In-Depth Analysis on HRSID, RSDD and SSDD Datasets. In Proceedings of the 2025 IEEE 4th International Conference on Computing and Machine Intelligence (ICMI), 2025; pp. 1-5.
42. Li, J.; Qu, C.; Shao, J. Ship detection in SAR images based on an improved faster R-CNN. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA), 2017; pp. 1-6.
43. Wang, Y.; Wang, C.; Zhang, H.; Dong, Y.; Wei, S.J.R.S. Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery. **2019**, *11*, 531.
44. Zhang, H.; Yu, H.; Tao, Y.; Zhu, W.; Zhang, K.J.I.I.P. Improvement of ship target detection algorithm for YOLOv7-tiny. **2024**, *18*, 1710-1718.
45. Luo, Y.; Li, M.; Wen, G.; Tan, Y.; Shi, C.J.I.A. SHIP-YOLO: A lightweight synthetic aperture radar ship detection model based on YOLOv8n algorithm. **2024**, *12*, 37030-37041.
46. Wu, C.-M.; Lei, J.; Liu, W.-K.; Ren, M.-L.; Ran, L.-L.J.C., Materials; Continua. Unmanned Ship Identification Based on Improved YOLOv8s Algorithm. **2024**, *78*.
47. He, L.-h.; Zhou, Y.-z.; Liu, L.; Cao, W.; Ma, J.-h.J.S.R. Research on object detection and recognition in remote sensing images based on YOLOv11. **2025**, *15*, 14032.
48. Han, Y.; Qi, K.; Zheng, J.; Li, J.; Jiang, F.; Zhang, X.; You, W.; Zhang, X.J.S.A. Lightweight cattle facial recognition method based on improved YOLOv11. **2025**, *7*, 173-184.

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.