ICJK
Chinese Journal of Information Fusion
http://dx.doi.org/10.62762/CJIF.2025.982112

RESEARCH ARTICLE

Lightweight SAR Ship Detection Network Based on Adaptive Spatial Feature Fusion and Channel Attention

Chi Zhang[1], **Shuyi Zhao**[1,*], **Jun Ma**[1], **Xunhuan Ren**[1] **and Viktar Yurevich Tsviatkou**[1]

[1] Department of Infocommunication Technologies, Belarusian State University of Informatics and Radioelectronics, Minsk 220013, Belarus

Abstract

Ship detection in Synthetic Aperture Radar (SAR) images remains challenging due to coherent speckle noise, complex inshore clutter, and large variations in target scale, especially for tiny ships. To address these issues, this paper proposes a lightweight SAR ship detection network based on YOLOv11n. The proposed model introduces a high-resolution P2 detection branch to preserve fine spatial details that may be weakened during repeated downsampling. To improve multi-scale feature representation, a Four-Head Adaptive Spatial Feature Fusion (FASFF) structure is adopted to adaptively combine features from P2, P3, P4, and P5. In addition, the Squeeze-and-Excitation (SE) attention module is inserted into the high-resolution branches of the Neck to recalibrate channel responses and suppress clutter-dominant features during feature aggregation. Experiments on a fixed subset of the SAR-Ship-Dataset show that the proposed YOLOv11n-SE-FASFF model improves Recall and mAP50-95 compared with the YOLOv11n baseline. Specifically, Recall increases from 0.898 to 0.926, and mAP50-95 increases from 0.614 to 0.632. The model contains 3.01 M parameters and achieves an inference speed of 196 FPS with 640 × 640 input images on an NVIDIA GeForce RTX 5060Ti GPU. Additional cross-dataset evaluation on SSDD and HRSID without fine-tuning further suggests that the proposed model has a certain degree of transferability under different SAR data conditions.

Keywords: SAR ship detection, YOLOv11, small target detection.

1 Introduction

Synthetic Aperture Radar (SAR), as an active microwave remote sensing technology, plays an irreplaceable role in maritime monitoring, maritime rescue, and illegal fishing control due to its capability of all-day and all-weather observation [1]. However, practical SAR ship detection in inshore scenes remains challenging. Coherent imaging introduces speckle noise, generating strong random textures that reduce target–background separability; meanwhile, man-made structures near coastlines and ports often produce strong backscatter responses similar to ship highlights, leading to false alarms. In addition, ship

Academic Editor:
Jungang Yang

Submitted: 29 December 2025
Accepted: 08 June 2026
Published: 13 June 2026

Vol. 3, **No.** 2, 2026.
10.62762/CJIF.2025.982112

***Corresponding author:**
Shuyi Zhao
2604925959@qq.com

Citation
Zhang, C., Zhao, S., Ma, J., Ren, X., & Tsviatkou, V. Y. (2026). Lightweight SAR Ship Detection Network Based on Adaptive Spatial Feature Fusion and Channel Attention. *Chinese Journal of Information Fusion*, 3(2), 138–152.

scales vary drastically and tiny vessels may occupy only a few pixels, making them particularly vulnerable to feature fading or loss under repeated downsampling and cross-scale fusion. Early SAR ship detection primarily relied on Constant False Alarm Rate (CFAR) algorithms and their variants. While effective in simple scenarios, these methods, which depend on statistical distribution assumptions, often struggle to balance false alarm rates and missed detection rates when facing complex sea states and multi-scale targets [2]. Rohling analyzed CFAR thresholding in clutter and multiple-target situations and showed that when interfering targets or strong scatterers enter the reference window, the estimated background level is biased upward, which raises the threshold and causes weaker targets to be masked (reduced detection probability) and leads to unstable behavior around clutter transitions and edges [3]. Gandhi and Kassam provided a systematic analysis of CFAR processors in nonhomogeneous backgrounds, demonstrating that when the assumption of identically distributed reference samples is violated, the constant-false-alarm property can be lost and the detector exhibits performance losses and distorted false-alarm control (often discussed in terms of CFAR loss and sensitivity to background mismatch) [4]. Moreover, sea clutter often exhibits heavy-tailed statistics rather than simple Gaussian behavior; the IET monograph on sea clutter summarizes scattering mechanisms and statistical modeling (including the K-distribution), clarifying why heavy-tailed backgrounds make threshold selection more sensitive and why extreme samples can disproportionately affect local estimates and hence threshold stability [5].

In recent years, with the rapid development of Convolutional Neural Networks (CNNs), deep learning-based object detection algorithms have gradually superseded traditional methods. SAR ship detection has gradually shifted from being "threshold and statistical hypothesis-driven" toward "data-driven representation learning." In the field of general object detection, two-stage and single-stage frameworks have established a mature ecosystem and have been extensively adapted for SAR tasks. Two-stage detectors, represented by Faster R-CNN, rely on a Region Proposal Network (RPN) to generate dense candidate regions on feature maps, followed by objectness scoring and coarse regression. Subsequently, a small subset of candidates undergoes RoI (Region of Interest) feature alignment and more robust classification/regression. While this paradigm often yields more stable discrimination and localization in complex backgrounds [6], its lightweight implementation and real-time deployment are constrained by the additional computational overhead of proposal generation and RoI-level processing.

In contrast, single-stage detectors — exemplified by SSD, RetinaNet, and the YOLO series–streamline the inference pipeline by performing dense classification and bounding box regression directly on multi-scale feature maps. SSD introduced the paradigm of multi-scale detection using default boxes across various layers, though its limited semantic information in shallow layers often hinders small object discrimination [7]. RetinaNet addressed the extreme foreground-background imbalance by proposing Focal Loss atop a Feature Pyramid Network (FPN), significantly enhancing single-stage accuracy to levels comparable with two-stage detectors [8].

The YOLO (You Only Look Once) lineage emphasizes end-to-end efficiency and industrial deployability. YOLOv3 [9] utilizes multi-scale detection heads (typically three scales) and incorporates FPN-like cross-layer fusion to bolster multi-scale detection capabilities. YOLOv4 [10] systematically integrated a suite of "plug-and-play" enhancements, categorized as "bag-of-freebies" and "bag-of-specials," to further optimize the speed-accuracy trade-off. More recently, YOLOv7 [11] advanced real-time performance through architectural optimizations, model re-parameterization, and refined training strategies. Given that SAR ship detection necessitates high recall, low false alarm rates, and deployable real-time inference, this study adopts a lightweight YOLO implementation as the engineering baseline. We further introduce targeted enhancements to the feature fusion modules and detection heads to address SAR-specific challenges, such as speckle noise, shoreline clutter, and the high risk of missing small targets. However, directly applying generic YOLO models to SAR images remains challenging. First, the inherent coherent speckle noise in SAR images is easily confused with tiny ships, leading to false alarms. Second, the drastic variation in ship scales and complex inshore backgrounds cause standard Feature Pyramid Networks (FPN) to lose feature information of tiny targets (e.g., smaller than 10 pixels) during down-sampling. This challenge is particularly evident in publicly available SAR ship benchmarks such as SSDD [12], where a significant proportion of ship instances are small-scale targets that are vulnerable

to feature fading under repeated downsampling and cross-scale fusion.

Although subsequent versions such as YOLOv8 [13] and YOLOv10 [14] have further improved detection efficiency, directly applying generic YOLO detectors to SAR ship images remains challenging because they are not specifically designed for coherent speckle noise, inshore clutter, and tiny ship targets. To address these issues, this paper proposes a lightweight SAR ship detection network based on YOLOv11n. The proposed model introduces a high-resolution P2 detection branch to preserve fine spatial details for tiny ships that may be weakened during repeated downsampling. A Four-Head Adaptive Spatial Feature Fusion (FASFF) [15] structure is adopted to adaptively combine features from P2, P3, P4, and P5, allowing the network to select more suitable scale information for ships of different sizes. In addition, the Squeeze-and-Excitation (SE) [16] module is inserted into the high-resolution branches of the Neck, where shallow features contain both useful tiny-ship cues and strong speckle or clutter responses. This design enables channel recalibration during feature aggregation and helps suppress clutter-dominant responses before final prediction.

The main contributions of this paper are summarized as follows:

- A lightweight YOLOv11n-based SAR ship detection network is developed by introducing a high-resolution P2 detection branch and a FASFF structure, which helps preserve fine-grained spatial cues for tiny ship targets during multi-scale feature fusion.
- The SE attention module is embedded into the high-resolution branches of the Neck. This placement is designed to recalibrate channel responses during feature aggregation and reduce clutter-dominant feature responses in complex inshore SAR scenes.
- Experiments on SAR-Ship-Dataset, together with cross-dataset evaluation on SSDD and HRSID, are conducted to evaluate the proposed method. The results show that the proposed model improves Recall and mAP50-95 compared with the YOLOv11n baseline while maintaining a lightweight structure.

2 Related Work

SAR ship detection in complex inshore environments is challenged by both scale variation and imaging noise. Ships exhibit large size diversity and tiny instances occupy only a few pixels, making fine-grained localization cues vulnerable to downsampling and cross-level fusion. Meanwhile, coherent SAR imaging introduces speckle noise, and strong scatterers from coastlines and harbors often produce ship-like responses, degrading target–background separability. To improve robustness, existing studies mainly progress along two directions: multi-scale representation enhancement and speckle clutter suppression.

Representative multi-scale approaches are commonly built upon pyramid architectures (e.g., FPN [17]/PAN [18] -style designs), where shallow high-resolution features provide spatial details and deeper layers provide stronger semantics. On top of this backbone, prior works primarily improve how multi-level features are propagated and fused, typically via attention mechanisms, scale-aware alignment, or adaptive weighting. One stream enhances pyramids with attention-guided feature reweighting. For instance, dense attention pyramid networks (DAPN [19]) typically introduce attention operations across pyramid levels — implemented as spatial attention, channel attention, or their combinations-so that ship-relevant regions/channels are emphasized during fusion, alleviating the tendency of tiny-ship cues to be overwhelmed by complex backgrounds. Another stream focuses on scale-aware modeling and scale alignment. Scale-aware pyramid networks such as SARFNet [20] usually address the mismatch between object scales and the effective receptive fields of different pyramid levels, using scale-dependent weighting, scale-equalizing convolutions, or alignment strategies to produce more consistent responses for ships of varying sizes. In addition, multi-scale feature extraction with adaptive fusion [21] often employs parallel branches (e.g., different kernel sizes, dilation rates, or multi-level features) and then merges them via gating or learned weights, aiming to improve scale robustness and boost recall on small targets.

While these methods improve cross-scale detection, their main focus remains on fusion design across feature levels. In SAR inshore settings, however, speckle and strong scatterers can dominate shallow textures and edge-like patterns; without an explicit mechanism to suppress textures or channels

dominated by noise or clutter during aggregation, stronger multi-scale fusion may inadvertently amplify nuisance responses as well, resulting in increased false alarms or unstable localization under severe clutter.

Recent studies on small and tiny object detection have also provided useful insights for SAR ship detection. Different from generic object detection, tiny object detection requires stronger preservation of high-resolution spatial cues and more flexible allocation of detection queries or feature responses. DQ-DETR [22] addresses tiny object detection from the perspective of dynamic query design. It introduces a categorical counting module, counting-guided feature enhancement, and dynamic query selection to adjust the number and positional information of object queries according to the object distribution in each image. This design is effective for aerial tiny object detection, where object instances are small and unevenly distributed. However, DETR-like models usually involve more complex training strategies and higher computational cost, which may limit their direct deployment in lightweight SAR ship detection scenarios.

SpecDETR [23] extends Transformer-based detection to hyperspectral point object detection. It treats each pixel as a token and uses a multi-layer Transformer encoder with self-excited subpixel-scale attention to extract spatial-spectral features for extremely small point-like targets. Although hyperspectral point object detection differs from SAR ship detection in data modality, it highlights the importance of subpixel-level feature enhancement and precise localization for very small targets. RT-DETRv3 [24] further improves real-time end-to-end detection by introducing hierarchical dense positive supervision to alleviate the sparse supervision problem of Hungarian matching in RT-DETR. These works indicate that recent small-object detection research is moving toward dynamic query allocation, dense supervision, and fine-grained feature enhancement. In contrast, the present work focuses on a lightweight CNN-based SAR detector. By introducing a high-resolution P2 branch and FASFF-based adaptive multi-scale fusion, the proposed method aims to preserve small-ship spatial details while maintaining efficient inference.

Speckle and clutter suppression for SAR ship detection is commonly developed along two closely related lines. A classic paradigm applies despeckling or clutter suppression as a preprocessing step and then runs a detector, where stronger smoothing typically reduces speckle more aggressively but may also attenuate weak scattering signatures and blur fine ship edges, especially for tiny targets in inshore scenes [25]. To avoid relying on explicit image-level smoothing, speckle-robust or "speckle-free" detectors instead construct local-window statistics that are less sensitive to speckle, such as homogeneity, consistency measures or similarity-based tests, and use these cues to enhance target–background separability under heavy clutter [26]. More recently, some works couple suppression with detection by integrating noise and clutter mitigation modules into the detection network, e.g., via attention-based feature recalibration that down-weights clutter-dominant channels and suppresses nuisance responses during representation learning.

Recent remote-sensing and SAR target detection studies have also explored more diverse learning paradigms. Xiao et al. [27] proposed a highly efficient and unsupervised framework for moving object detection in satellite videos, where pseudo labels generated by traditional methods are progressively refined during training and sparse spatio-temporal representation is used to reduce redundant background computation. Although this method is designed for satellite video moving-object detection rather than SAR ship detection, it highlights the importance of annotation efficiency and computational efficiency for weak and small targets in remote-sensing scenarios. MaDiNet [28] introduces a Mamba diffusion network for SAR target detection, where diffusion-based modeling and the MambaSAR module are used to capture spatial structural information and distinguish targets from complex SAR backgrounds. These recent studies show that remote-sensing and SAR target detection is moving toward more robust representation learning under complex backgrounds and limited annotations.

Overall, existing SAR ship detection and small-object detection methods have demonstrated the importance of multi-scale representation, attention-based feature refinement, fine-grained localization, and robust representation learning under complex backgrounds. However, many recent Transformer-, diffusion-, or Mamba-based detectors introduce relatively high computational cost and usually require carefully tuned training strategies. For practical SAR ship monitoring, especially in resource-constrained scenarios, a lightweight detector with strong small-target sensitivity and clutter robustness is still

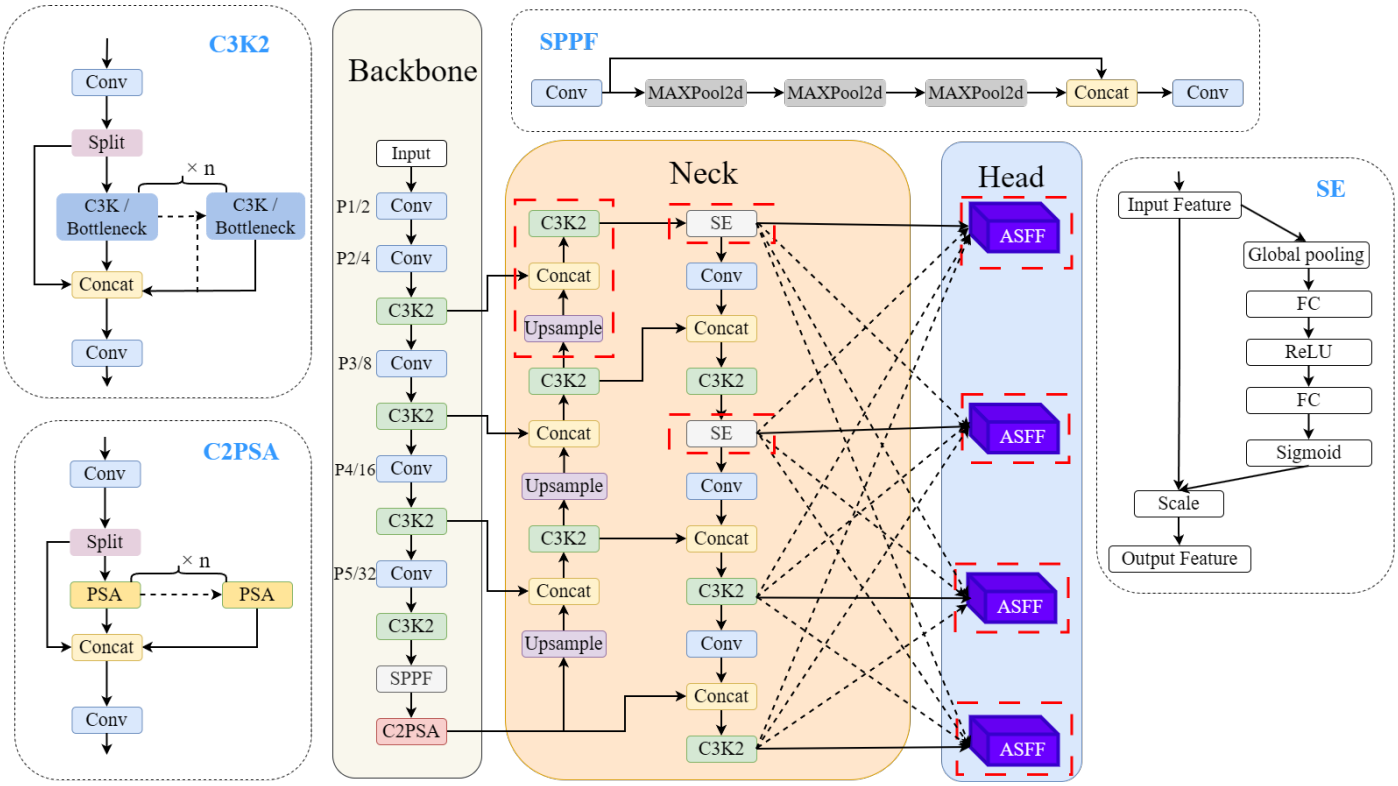

Figure 1. YOLOv11-SE-FASFF Architecture Overview. The model consists of three main parts: Backbone, Neck, and Head. The Backbone extracts multi-level features from the input SAR image, while the Neck performs hierarchical feature aggregation. SE attention modules are inserted into the high-resolution branches of the Neck to recalibrate channel responses and suppress clutter-dominant features. The Head adopts the FASFF structure to adaptively fuse P2, P3, P4, and P5 features before prediction, improving the representation of tiny and multi-scale ship targets.

necessary. Motivated by these observations, this paper improves YOLOv11n by introducing a high-resolution P2 branch, FASFF-based adaptive four-scale fusion, and SE-based channel recalibration in the Neck, aiming to balance detection accuracy, noise robustness, and inference efficiency.

3 Methodology

To address the challenges of severe coherent speckle noise and the difficulty of detecting multi-scale ships (especially tiny ones) in inshore SAR scenarios, this paper proposes an improved one-stage detection network. This section first outlines the overall network architecture, followed by a detailed elaboration of the two core improvement modules: the Squeeze-and-Excitation (SE) attention mechanism for noise suppression and the Four-Head Adaptive Spatial Feature Fusion (FASFF) strategy for resolving multi-scale feature conflicts.

3.1 Overall Network Architecture

The proposed model is built upon the YOLOv11n architecture. To adapt to the characteristics of SAR images, which feature complex inshore backgrounds and numerous tiny ships, we reconstructed the original architecture as shown in Figure 1.

The improved network consists of three main components:

1. Backbone: Responsible for feature extraction. Unlike the original YOLOv11, we utilize shallower features to preserve texture information of tiny objects. Consequently, the backbone outputs feature maps at four scales (P2, P3, P4, P5). Specifically, the P2 layer contains rich high-resolution details, which are crucial for capturing tiny ships.

2. Neck: A hierarchical feature fusion structure is employed for multi-scale feature aggregation. Considering that coherent speckle noise in SAR images is predominantly concentrated in high-frequency texture regions, we embedded Squeeze-and-Excitation (SE) attention modules specifically at the output of the high-resolution P2 and P3 feature layers to precisely suppress background clutter. The P4 and P5 layers, which contain richer semantic information, remain unchanged to ensure efficient information flow.

3. Head: We introduce a Four-Head Adaptive Spatial Feature Fusion (FASFF) detection head. This module receives features from P2, P3, P4, and P5. It utilizes adaptive weights to resolve feature conflicts across the large scale span (from the tiny-scale P2 to the large-scale P5), finally generating the detection results.

3.2 Squeeze-and-Excitation Attention Mechanism

In the SAR image ship detection task, in order to deal with complex background interference (such as coastline, wave texture) and inherent coherent speckle noise, this paper integrates the Squeeze-and-Excitation (SE) attention module [16] to the neck of YOLOv11, aiming to enhance the model's sensitivity to effective features and suppress useless background. The structure of the Squeeze-and-Excitation (SE) attention module is shown in Figure 2. In image detection, small targets often rely on shallow high-resolution features for location, but these feature layers (P2 and P3) are also filled with the most severe coherent speckle noise and background clutter, and direct fusion often leads to an increase in false alarm rates. To this end, this paper does not place the attention mechanism in the deep backbone network, but specifically embeds it into the P2 and P3 high-resolution branches of the YOLOv11 Neck feature fusion network.

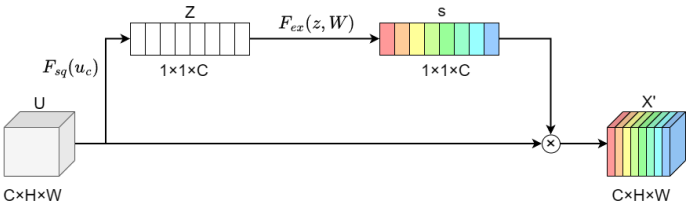

Figure 2. Schematic diagram of the SE attention module. The module first summarizes the input feature map into a compact channel descriptor and then learns channel-wise importance weights. These weights are applied to the original feature map through feature scaling, producing a recalibrated output feature map. This mechanism helps the network emphasize informative ship-related channels while reducing the influence of background clutter and speckle-related responses.

The SE module adaptively recalibrates the feature response by explicitly modeling the interdependence between feature channels.

The core process of the SE module includes two stages: Squeeze and Excitation. The Squeeze stage compresses the spatial feature map of spatial dimension $H \times W$ into a 1×1 channel descriptor Z_c through Global Average Pooling. The Excitation stage uses two fully connected layers (FC) to learn the nonlinear relationship between channels and generate the channel weight vector s. Their mathematical formulations are systematically presented in the following.

$$Z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j), \tag{1}$$

$$s = F_{ex}(z, W) = \sigma(W_2 \delta(W_1 z)), \tag{2}$$

where Z_c denotes the squeezed global statistic for the c-th channel, and $u_c(i,j)$ represents the input feature value at spatial position (i,j) with dimensions $H \times W$. The terms δ and σ refer to the ReLU and Sigmoid activation functions, respectively. W_1 and W_2 are the learned weight matrices of the two fully connected layers (dimensionality reduction and expansion), and s is the final adaptive weight vector used to recalibrate the features.

3.3 Detection Head with Adaptive Spatial Feature Fusion

Although YOLOv11 employs an advanced PANet structure for multi-scale feature fusion, it still faces the challenge of "scale inconsistency" when dealing with ship targets exhibiting extreme scale variations in SAR images. Specifically, features at different levels of the pyramid contain contradictory information: deep layers (e.g., P5) possess strong semantics but lack geometric details, while shallow layers (e.g., P2) retain details but are filled with speckle noise. Traditional fusion methods, such as summation or concatenation, force the mixing of background noise with target features, often causing tiny ships to be overwhelmed or leading to false alarms.

To address this limitation, we introduced the Four-Head Adaptive Spatial Feature Fusion (FASFF) mechanism prior to the four detection heads (P2, P3, P4, P5) of YOLOv11 as shown in Figure 3. The core philosophy of FASFF is to spatially filter conflicting information, enabling the network to autonomously learn the importance weights of features from different scales at each spatial location, thereby preserving only the informative features for fusion.

The proposed FASFF module is designed to perform adaptive feature selection before the final detection heads. Unlike conventional feature fusion operations such as direct summation or concatenation, FASFF does not assign a fixed contribution to each pyramid level. Instead, for each detection scale, features from

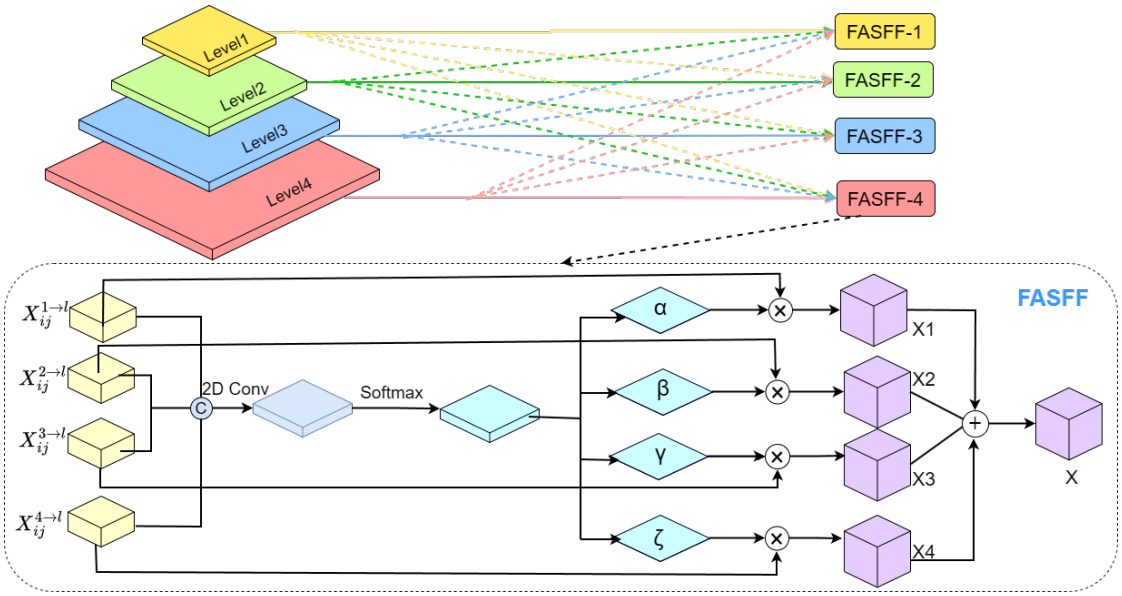

Figure 3. Structure of the proposed FASFF module. For each detection level, feature maps from P2, P3, P4 and P5 are first resized to the same spatial resolution. A lightweight convolutional layer predicts spatial fusion weights, which are normalized by Softmax. The final output feature is generated by weighted summation and then sent to the corresponding detection head.

P2, P3, P4 and P5 are first resized to the same spatial resolution. Then, a lightweight convolutional layer is used to predict spatially varying fusion weights for the four feature levels. These weights are normalized by a Softmax operation and then used to generate the fused feature map through weighted summation.

In this way, the network can adaptively determine which scale should contribute more at different spatial locations. For example, regions containing tiny ships can rely more on high-resolution P2 or P3 features, while regions containing larger vessels or complex structures can receive more semantic information from P4 or P5. This design is particularly suitable for SAR ship detection, where target sizes vary greatly and shallow features may contain both useful boundary details and strong speckle noise.

Let X^l denote the feature map at level l (l in 1, 2, 3, 4, corresponding to P2 through P5). To fuse features for level l, we first resize the features X^n from all other levels n ($n \neq l$) to the same resolution as X^l via up-sampling or down-sampling, denoted as $X^{n\to l}$. Subsequently, 1×1 convolutional layers are employed to predict four weight parameters α_{ij}^l, β_{ij}^l, γ_{ij}^l, ζ_{ij}^l at spatial location (i, j). To ensure the weights are normalized, a Softmax function is applied:

$$\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l + \zeta_{ij}^l = 1, \tag{3}$$

$$\alpha_{ij}^l = \frac{e^{\lambda_{ij}^l}}{e^{\lambda_{ij}^l} + e^{\mu_{ij}^l} + e^{\nu_{ij}^l} + e^{\zeta_{ij}^l}}, \tag{4}$$

where λ, μ, ν, ζ represent the control parameters output by the convolution layers.

Finally, the fused feature Y^l generated by the FASFF module is calculated as:

$$Y_{ij}^l = \alpha_{ij}^l \cdot X_{ij}^{1\to l} + \beta_{ij}^l \cdot X_{ij}^{2\to l} + \gamma_{ij}^l \cdot X_{ij}^{3\to l} + \zeta_{ij}^l \cdot X_{ij}^{4\to l}, \tag{5}$$

Through pixel-wise adaptive weighting, FASFF allows the model to select different feature levels according to local image content. When a small ship appears in a high-resolution region, the fusion process can assign relatively larger weights to P2 or P3 features, thereby preserving fine spatial cues that may be weakened in deeper layers. When the target is larger or the surrounding background is more complex, deeper features from P4 or P5 can provide stronger semantic guidance. Therefore, FASFF does not simply increase the number of detection heads; instead, it enables scale-aware feature selection across the four pyramid levels. This helps reduce the feature conflict between shallow noisy details and deep semantic representations, improving the robustness of multi-scale SAR ship detection.

4 Experiments

4.1 Dataset and Implementation Details

The experiments were mainly conducted on SAR-Ship-Dataset [29], a public SAR ship detection dataset containing 43,819 SAR images with a resolution of 640 × 640 pixels. Considering the focus of this work on lightweight architectural comparison and repeated ablation experiments, we used a fixed representative subset from the official dataset. Specifically, 1,380 images were selected from the official training set for model training, and 361 images were selected from the official test set for evaluation. The resulting train-test ratio is approximately 3.8:1, which is close to the commonly used 80%/20% setting. The same training and test subsets were used for all compared models to ensure a fair and controlled comparison.

The subset-based protocol was adopted for two reasons. First, the main objective of this study is to evaluate whether the proposed SE placement and FASFF structure can bring consistent improvements under identical experimental conditions, rather than to report the highest possible benchmark result using all available data. Second, repeated training of multiple baseline, ablation, and comparison models on the full dataset would require substantially higher computational cost. To reduce sampling bias, the selected images were kept fixed throughout all experiments and included both inshore and offshore scenes with different target scales and background conditions. All experiments were implemented within the PyTorch framework. The input image size was fixed to 640 × 640 for both training and inference. To ensure a fair comparison among YOLO-based models, YOLOv11n, YOLOv8n, YOLOv10n, and the proposed YOLOv11n-SE-FASFF model used the same dataset split, input size, optimizer, training epochs, and augmentation settings. The optimizer employed was Stochastic Gradient Descent (SGD) with an initial learning rate of 0.01, a momentum of 0.937, and a weight decay of 0.0005. Due to GPU memory constraints, for the YOLO-based models, the batch size was set to 12, and the models were trained for 300 epochs. During training, standard YOLO-style data augmentation was used, and mosaic augmentation was enabled with a probability of 1.0. The same augmentation setting was applied to the YOLO-based models. Performance was evaluated using standard metrics, including Precision (P), Recall (R), mAP50, and mAP50-95, to provide a holistic assessment of detection accuracy and localization capability. Their mathematical formulations are systematically presented in the following.

$$P = \frac{TP}{TP + FP}, \tag{6}$$

where TP is true positive, FP is false positive.

$$R = \frac{TP}{TP + FN}, \tag{7}$$

where FN is false negative.

$$mAP_{50} = \frac{1}{C}\sum_{c=1}^{C} AP_c\,(IoU = 0.5)\,, \tag{8}$$

where C is object categories, and AP_c is average precision for class c.

$$mAP_{50-95} = \frac{1}{10}\sum_{k=0}^{9} mAP_{0.5+0.05k}, \tag{9}$$

where $mAP50$ denotes the mean Average Precision calculated at an IoU threshold of 0.50, while $mAP50-95$ denotes the mean Average Precision averaged over IoU thresholds from 0.50 to 0.95 with a step size of 0.05. The $mAP50$ metric reflects detection performance under a relatively loose localization criterion, whereas $mAP50-95$ provides a stricter evaluation of localization accuracy across multiple IoU thresholds.

To ensure reproducibility, the experiments were performed on a single laptop computer equipped with an NVIDIA GeForce RTX 5060Ti GPU, using Python 3.9.24 and PyTorch 2.8.0.

4.2 Ablation Study Analysis

To investigate the specific contributions of the proposed modules and the influence of SE placement, a stepwise ablation study was conducted based on the YOLOv11n baseline, with results presented in Table 1.

As shown in Table 1, the baseline YOLOv11n achieves a Precision of 0.927, Recall of 0.898, mAP50 of 0.967, and mAP50-95 of 0.614. After introducing the FASFF structure, the Recall increases from 0.898 to 0.916, while mAP50-95 improves from 0.614 to 0.626. This result indicates that the high-resolution P2 detection branch and adaptive spatial feature fusion are helpful for reducing missed detections. In SAR images, tiny ships may occupy only a few pixels and can

Table 1. Ablation study of the proposed modules and SE placement.

Methods	P_ψ	R_ψ	mAP50	mAP50-95
YOLOv11n	0.927	0.898	0.967	0.614
YOLOv11n+FASFF	0.930	0.916	0.967	0.626
YOLOv11n+SE (Backbone)	0.908	0.915	0.966	0.607
YOLOv11n+SE (Neck)	**0.938**	0.893	0.967	0.627
YOLOv11n+SE (Neck)+FASFF	0.932	**0.926**	**0.971**	**0.632**

be weakened during repeated downsampling. By adaptively fusing P2, P3, P4, and P5 features, FASFF helps preserve fine-grained spatial cues for small targets.

The influence of SE placement was further examined by comparing Backbone and Neck insertion strategies. When SE is inserted into the Backbone, Recall increases from 0.898 to 0.915, but Precision decreases from 0.927 to 0.908, and mAP50-95 decreases from 0.614 to 0.607. This suggests that applying channel recalibration at an early feature extraction stage may enhance more candidate responses, including some clutter-like regions, but it does not effectively improve localization stability. In contrast, placing SE in the Neck achieves the highest Precision of 0.938 and improves mAP50-95 from 0.614 to 0.627. It is noted that SE (Neck) placement results in a slight decrease in Recall from 0.898 to 0.893 compared with the baseline, suggesting that channel selectivity in the Neck may occasionally suppress weak-scattering ship responses that are similar in appearance to clutter. This result supports the design choice of inserting SE into the high-resolution Neck branches. The P2 and P3 branches retain fine spatial details for tiny ships, but they are also more sensitive to speckle noise and inshore clutter. Applying SE at this stage allows the network to recalibrate channel responses after initial feature extraction and before final prediction, which is more suitable for suppressing clutter-dominant features while retaining useful ship-related information.

It should be noted that SE placement in the Backbone obtains a higher Recall than SE placement in the Neck, but its lower Precision and mAP50-95 indicate more false responses or less stable localization. Therefore, Neck placement provides a more balanced choice for clutter suppression and localization accuracy. In this work, SE was not placed immediately before the detection head because the noisy shallow features have already been propagated through the feature fusion pathway at that stage. Although placing SE immediately before the detection head was also considered, this configuration was not included in the ablation study because the shallow noisy features have already been propagated through the feature fusion pathway at that stage, making Neck-level insertion a more principled design choice. In contrast, inserting SE into the high-resolution Neck branches allows channel recalibration to be performed during feature aggregation, which helps suppress clutter-dominant responses before they are sent to the final prediction layers.

The full YOLOv11n-SE-FASFF model further combines the advantages of SE and FASFF, achieving a Precision of 0.932, Recall of 0.926, mAP50 of 0.971, and mAP50-95 of 0.632. Compared with the baseline, Recall increases by 2.8 percentage points and mAP50-95 increases by 1.8 percentage points. These results indicate that SE and FASFF play complementary roles in SAR ship detection: FASFF improves the sensitivity to small and multi-scale ship targets, while SE helps reduce clutter-related false responses.

Although the FASFF structure and SE attention modules introduce additional operations, the proposed model still maintains a lightweight design. As shown in the evaluation, the parameter count increases from 2.59 M in the baseline YOLOv11n to 3.01 M in the proposed model, corresponding to an increase of only 0.42 M. The inference speed was measured using 640 × 640 input images on a single NVIDIA GeForce RTX 5060Ti GPU, and the proposed model achieved 196 FPS under this setting. The threshold of 30 FPS was adopted as a practical reference for real-time processing, because it corresponds to approximately 33 ms per frame. Although SAR ship detection is not always a video-stream task, this threshold provides an engineering reference for near-real-time maritime monitoring applications. Overall, these results suggest that the proposed model improves detection performance while introducing only a limited increase

Table 2. Comparison of the performance metrics of different models based on the SAR-Ship-Dataset.

Methods	P_c	R_c	mAP50	mAP50-95
YOLOv11n	0.927	0.898	0.967	0.614
YOLOv8n	0.930	0.911	**0.971**	0.620
YOLOv10n	0.883	0.891	0.949	0.606
RT-DETR-l	0.860	0.858	0.925	0.540
YOLOv11n+SE+FASFF	**0.932**	**0.926**	**0.971**	**0.632**

in computational cost.

4.3 Comparative Analysis

To further evaluate the effectiveness of the proposed algorithm in SAR ship detection tasks, we conducted comparative experiments with several representative detection models, including YOLOv8n [13], YOLOv10n [14], and RT-DETR-l. The YOLO-based models were trained using the same dataset split, input image size, and main training settings. Due to the higher memory consumption of RT-DETR-l, its training batch size was set to 2, while the YOLO-based models were trained with a batch size of 12. It should be noted that the different batch sizes may affect optimization dynamics and BatchNorm statistics, which could partially contribute to the performance gap between RT-DETR-l and the YOLO-based models under the current experimental setting. All models were evaluated on the same test subset with an input size of 640 × 640. The detailed comparison results are presented in Table 2.

As shown in Table 2, the proposed model achieves the highest Recall and mAP50-95 among the compared models. Compared with the baseline YOLOv11n, the Recall increases from 0.898 to 0.926, with a gain of 2.8 percentage points. This result suggests that the introduced P2 high-resolution detection branch and FASFF structure are helpful for reducing missed detections of tiny ship targets, which are easily weakened during deep feature propagation.

In the comparison with other lightweight YOLO detectors, YOLOv8n shows competitive performance, achieving a mAP50 of 0.971. However, its mAP50-95 is 0.620, which is slightly lower than that of the proposed method. This indicates that YOLOv8n can detect most ship targets under a relatively loose IoU threshold, but its localization accuracy under stricter IoU thresholds is still affected by complex SAR backgrounds.

YOLOv10n obtains a lower mAP50-95 of 0.606 on the SAR-Ship-Dataset. One possible reason is related to its NMS-free detection design. Although the NMS-free strategy improves inference efficiency by reducing the dependence on post-processing, it also requires the model to learn more accurate confidence ranking, duplicate suppression, and localization consistency during training. In SAR images, coherent speckle noise, weak scattering responses of tiny ships, and ship-like interference from coastlines, harbors, and artificial structures make the distinction between true targets and clutter more ambiguous than in ordinary optical images. Under such conditions, YOLOv10n may be more sensitive to noisy responses and localization uncertainty.

This effect is more evident in mAP50-95, because this metric evaluates detection performance under multiple IoU thresholds from 0.50 to 0.95. For tiny SAR ship targets, even a small bounding-box deviation may lead to a clear performance drop at high IoU thresholds. In contrast, the proposed model introduces SE-based channel recalibration and FASFF-based adaptive multi-scale fusion, which helps suppress clutter-dominant responses and preserve high-resolution spatial cues for small ships. This may explain why the proposed method achieves a higher mAP50-95 than YOLOv10n in this experiment.

RT-DETR-l was also included as a representative Transformer-based detector. Compared with the YOLO-based lightweight models, RT-DETR-l obtains lower values on all evaluation metrics, with a Precision of 0.860, Recall of 0.858, mAP50 of 0.925, and mAP50-95 of 0.540. This result may be related to its higher model complexity and stronger dependence on sufficient training data and carefully tuned optimization settings. In this study, the training subset contains 1,380 SAR images, which may be relatively limited for a large Transformer-based detector. In addition, the complex characteristics of SAR images, including speckle noise, weak target scattering, and inshore clutter, may increase the difficulty of learning stable object queries and accurate bounding-box localization. The smaller feasible batch size of RT-DETR-l may also affect optimization stability under

Table 3. Cross-dataset evaluation results and SAR-Ship-Dataset reference results.

Dataset	P_v	R_v	mAP50	mAP50-95
SSDD	0.967	0.958	0.985	0.732
HRSID	0.961	0.879	0.951	0.751
SAR-Ship-Dataset	0.932	0.926	0.971	0.632

the current hardware condition.

Overall, the proposed YOLOv11n-SE-FASFF model achieves a Precision of 0.932, Recall of 0.926, and the highest mAP50-95 of 0.632 among the compared models. Benefiting from SE-based channel recalibration and FASFF-based adaptive multi-scale fusion, the model improves small-target detection and localization stability while maintaining a lightweight structure. These results suggest that the proposed design provides a more balanced accuracy-efficiency trade-off for SAR ship detection under the current experimental setting.

4.4 Cross-Dataset Generalization Evaluation

To further examine the transferability of the proposed model, we conducted cross-dataset evaluation on SSDD [12] and HRSID [30]. In this experiment, the model was trained only on the selected training subset of SAR-Ship-Dataset and was directly tested on SSDD and HRSID without fine-tuning. Therefore, the results on SSDD and HRSID reflect cross-dataset generalization rather than in-domain training performance. The detailed test results are presented in Table 3.

SSDD is a widely used public benchmark dataset for SAR ship detection, containing SAR ship samples with different resolutions, polarization modes, and imaging conditions. As shown in Table 3, the proposed model achieves a Precision of 0.967, Recall of 0.958, mAP50 of 0.985, and mAP50-95 of 0.732 on SSDD. This result suggests that the features learned from SAR-Ship-Dataset can still be transferred to conventional SAR ship detection scenes to a certain extent.

HRSID is constructed from Sentinel-1B images and contains more complex high-resolution SAR scenes. When directly tested on HRSID without fine-tuning, the proposed model achieves a Precision of 0.961, Recall of 0.879, mAP50 of 0.951, and mAP50-95 of 0.751. Compared with SSDD, the Recall on HRSID is lower, which may be related to differences in image resolution, background complexity, target size distribution, and sensor characteristics. Nevertheless, the relatively high mAP50 and mAP50-95 indicate that the model still maintains stable localization performance under this cross-dataset setting.

Overall, the results on SSDD and HRSID suggest that the proposed YOLOv11n-SE-FASFF model has a certain degree of cross-dataset transferability. However, since the model was not fine-tuned on SSDD or HRSID, these results should be interpreted as evidence of transferability rather than direct in-domain benchmark comparison. Broader validation on more SAR datasets and sensor conditions is still needed for practical deployment.

4.5 Visualization Analysis

To provide an intuitive comparison of the detection results, several representative SAR images from the SAR-Ship-Dataset were selected for visualization, as shown in Figure 4. Column (a) presents the ground-truth annotations, while columns (b), (c), (d), and (e) show the results of YOLOv11n, YOLOv8n, YOLOv10n, and the proposed YOLOv11n-SE-FASFF model, respectively. In the figure, blue boxes denote detected ship targets, green boxes denote false detections, and red boxes denote missed ship targets.

For the baseline YOLOv11n, some false detections appear in regions with strong background scattering. This may be caused by the similarity between ship responses and speckle-like or artifact-like background patterns in SAR images. In addition, several tiny ships with weak scattering responses are not detected, indicating that shallow spatial details may be weakened during feature downsampling and fusion.

YOLOv8n shows competitive detection results, but some false alarms and missed detections can still be observed in complex background regions. YOLOv10n also detects a number of ship targets, but its localization and tiny-target detection performance appear less stable in the displayed examples. This is consistent with the quantitative results in Table 2, where YOLOv10n obtains a lower mAP50-95 than the other YOLO-based models.

Compared with these baseline models, the proposed

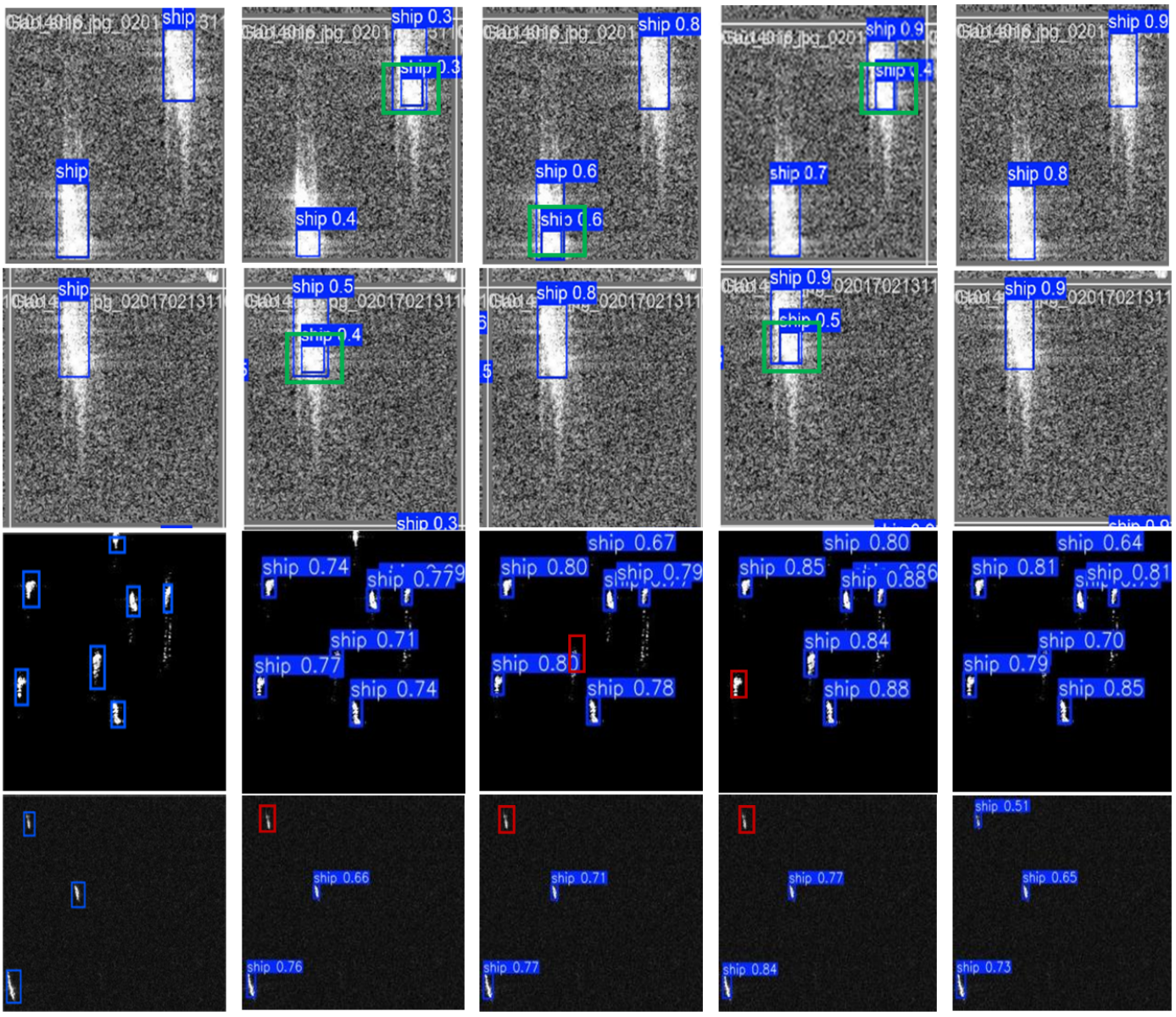

Figure 4. Visual comparison of SAR ship detection results on representative samples from the SAR-Ship-Dataset. Column (a) presents the ground-truth annotations, and columns (b), (c), (d), and (e) present the detection results of YOLOv11n, YOLOv8n, YOLOv10n, and the proposed YOLOv11n-SE-FASFF model, respectively. Blue boxes denote detected ship targets, green boxes denote false detections caused by speckle noise, artifacts, or ship-like background clutter, and red boxes denote missed ship targets.

YOLOv11n-SE-FASFF model shows more stable detection results in the displayed SAR scenes. The SE module helps reduce some clutter-related false responses, while the P2 branch and FASFF structure help preserve high-resolution spatial cues for tiny ships. As a result, fewer missed detections and more stable localization can be observed in these examples. These visualization results provide qualitative support for the quantitative improvements reported in the ablation and comparative experiments.

4.6 Limitations and Discussion

Although the proposed YOLOv11n-SE-FASFF model improves small-ship detection and clutter suppression under the current experimental setting, several limitations still need to be considered. First, SAR inshore scenes often contain dense artificial structures, docks, and strong scatterers whose backscattering responses may be similar to real ship targets. In such cases, false alarms may still occur because the visual difference between ship targets and ship-like clutter is relatively weak in SAR images.

Second, very small ships with weak scattering responses remain difficult to detect, especially when they are close to shoreline structures, strong sea clutter, or other bright scattering regions. Although the introduced P2 branch helps preserve high-resolution spatial details, extremely weak targets may still be suppressed or confused with noise-dominant background features.

Third, the current experiments are conducted mainly on a fixed subset of SAR-Ship-Dataset, with additional cross-dataset evaluation on SSDD and HRSID without fine-tuning. Therefore, the robustness of the proposed method under broader SAR imaging conditions, such as different sensors, resolutions, incidence angles, polarization modes, and sea states, still requires further validation.

These limitations suggest that future improvements should focus on more explicit speckle-aware feature modeling, stronger discrimination between ships and harbor clutter, and domain adaptation strategies for cross-sensor SAR ship detection.

5 Conclusion

This paper proposed a lightweight SAR ship detection network based on YOLOv11n for inshore SAR scenarios with speckle noise, complex clutter, and tiny ship targets. The proposed model combines a high-resolution P2 detection branch, FASFF-based adaptive four-scale feature fusion, and SE-based channel recalibration in the Neck. These designs help preserve fine spatial details for small ships, improve multi-scale feature selection, and suppress clutter-dominant channel responses during feature aggregation.

Experimental results indicate that the proposed YOLOv11n-SE-FASFF model improves Recall and mAP50-95 compared with the YOLOv11n baseline while maintaining a lightweight structure. The ablation study further shows that FASFF mainly contributes to small-target recall, while Neck-based SE placement provides better false-alarm suppression and localization stability than Backbone placement under the current experimental setting.

Future work will focus on improving robustness in dense harbor scenes, reducing missed detections of weak-scattering tiny ships, and validating the method on larger-scale SAR datasets with different sensors and imaging conditions. In addition, inspired by recent efficient unsupervised remote-sensing detection studies, weakly supervised or unsupervised learning strategies may be explored to reduce the dependence on manually annotated SAR data. More explicit speckle-aware feature modeling, cross-sensor domain adaptation, and further deployment optimization will also be investigated for practical onboard SAR ship detection.

Data Availability Statement

The source code supporting the findings of this study is publicly available at: https://github.com/cz1355303-afk/zc. The datasets used in this study are publicly available from their original repositories, including SAR-Ship-Dataset, SSDD, and HRSID. Additional data or materials that support the findings of this study are available from the corresponding author upon reasonable request.

Funding

This work was supported by the China Scholarship Council.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., & Papathanassiou, K. P. (2013). A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine*, *1*(1), 6-43. [CrossRef]

[2] Leng, X., Ji, K., Yang, K., & Zou, H. (2015). A bilateral CFAR algorithm for ship detection in SAR images. *IEEE Geoscience and Remote Sensing Letters*, *12*(7), 1536-1540. [CrossRef]

[3] Rohling, H. (1983). Radar CFAR thresholding in clutter and multiple target situations. *IEEE transactions on aerospace and electronic systems*, (4), 608-621. [CrossRef]

[4] Gandhi, P. P., & Kassam, S. A. (1988). Analysis of CFAR processors in nonhomogeneous background. *IEEE Transactions on Aerospace and Electronic systems*, *24*(4), 427-445. [CrossRef]

[5] Ward, K. D., Watts, S., & Tough, R. J. (2006). Sea clutter: scattering, the K distribution and radar performance (Vol. 20). IET. [CrossRef]

[6] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, 28, (pp. 91–99).

[7] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, September). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Cham: Springer International Publishing. [CrossRef]

[8] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2018). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(2), 318-327. [CrossRef]

[9] Redmon, J. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*. [CrossRef]

[10] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*. [CrossRef]

[11] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023, June). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7464-7475). IEEE. [CrossRef]

[12] Zhang, T., Zhang, X., Li, J., Xu, X., Wang, B., Zhan, X., ... & Wei, S. (2021). SAR ship detection dataset (SSDD): Official release and comprehensive data analysis. *Remote Sensing, 13*(18), 3690. [CrossRef]

[13] Terven, J., Córdova-Esparza, D. M., & Romero-González, J. A. (2023). A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine learning and knowledge extraction, 5*(4), 1680-1716. [CrossRef]

[14] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. *Advances in neural information processing systems, 37*, 107984-108011.

[15] Liu, S., Huang, D., & Wang, Y. (2019). Learning spatial fusion for single-shot object detection. *arXiv preprint arXiv:1911.09516*. [CrossRef]

[16] Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2019). Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(8), 2011-2023. [CrossRef]

[17] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017, July). Feature Pyramid Networks for Object Detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 936-944). IEEE. [CrossRef]

[18] Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018, June). Path Aggregation Network for Instance Segmentation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8759-8768). IEEE. [CrossRef]

[19] Cui, Z., Li, Q., Cao, Z., & Liu, N. (2019). Dense attention pyramid networks for multi-scale ship detection in SAR images. *IEEE Transactions on Geoscience and Remote Sensing, 57*(11), 8983-8997. [CrossRef]

[20] Tang, L., Tang, W., Qu, X., Han, Y., Wang, W., & Zhao, B. (2022). A scale-aware pyramid network for multi-scale object detection in SAR images. *Remote Sensing, 14*(4), 973. [CrossRef]

[21] Zhou, K., Zhang, M., Wang, H., & Tan, J. (2022). Ship detection in SAR images based on multi-scale feature extraction and adaptive feature fusion. *Remote Sensing, 14*(3), 755. [CrossRef]

[22] Huang, Y. X., Liu, H. I., Shuai, H. H., & Cheng, W. H. (2024, September). Dq-detr: Detr with dynamic query for tiny object detection. In *European Conference on Computer Vision* (pp. 290-305). Cham: Springer Nature Switzerland. [CrossRef]

[23] Li, Z., An, W., Guo, G., Wang, L., Wang, Y., & Lin, Z. (2025). SpecDETR: A transformer-based hyperspectral point object detection network. *ISPRS Journal of Photogrammetry and Remote Sensing, 226*, 221-246. [CrossRef]

[24] Wang, S., Xia, C., Lv, F., & Shi, Y. (2025, February). RT-DETRv3: Real-Time End-to-End Object Detection with Hierarchical Dense Positive Supervision. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 1628-1636). IEEE. [CrossRef]

[25] Huang, S. Q., Liu, D. Z., Gao, G. Q., & Guo, X. J. (2009). A novel method for speckle noise reduction and ship target detection in SAR images. *Pattern Recognition, 42*(7), 1533-1542. [CrossRef]

[26] Chen, S. W., Cui, X. C., Wang, X. S., & Xiao, S. P. (2021). Speckle-free SAR image ship detection. *IEEE Transactions on Image Processing, 30*, 5969-5983. [CrossRef]

[27] Xiao, C., An, W., Zhang, Y., Su, Z., Li, M., Sheng, W., ... & Liu, L. (2024). Highly efficient and unsupervised framework for moving object detection in satellite videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 46*(12), 11532-11539. [CrossRef]

[28] Zhou, J., Xiao, C., Peng, B., Liu, T., Liu, Z., Liu, Y., & Liu, L. (2025). MaDiNet: Mamba diffusion network for SAR target detection. *IEEE Transactions on Circuits and Systems for Video Technology, 35*(11), 10787–10800. [CrossRef]

[29] Wang, Y., Wang, C., Zhang, H., Dong, Y., & Wei, S. (2019). A SAR dataset of ship detection for deep learning under complex backgrounds. *Remote Sensing, 11*(7), 765. [CrossRef]

[30] Wei, S., Zeng, X., Qu, Q., Wang, M., Su, H., & Shi, J. (2020). HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation. *IEEE Access, 8*, 120234–120254. [CrossRef]

Chi Zhang is currently an undergraduate student at the Belarusian State University of Information and Radio Technology, where his research interests include image processing and machine learning. This work was supported by the China Scholarship Council. (Email: 3123150387@qq.com)

Shuyi Zhao is currently an undergraduate student at the Belarusian State University of Information and Radio Technology, where her research interests include image processing and machine learning. This work was supported by the China Scholarship Council. (Email: 2604925959@qq.com)

Jun Ma received the B.S. degree from the Lanzhou University of Technology, in 2015, and the M.S. degree from the Belarusian State University of Informatics and Radioelectronics, in 2018, where he is currently pursuing the Ph.D. degree. His current research interests include image processing, machine learning, and computer vision. (Email: majun@bsuir.by)

Xunhuan Ren received the B.S. degree from the Lanzhou University of Technology, in 2015, M.S. degree and Ph.D. degree from the Belarusian State University of Informatics and Radioelectronics, in 2018 and 2024, respectively, where she is currently an associate professor of the department of infocommunication technologies. Her current research interests include image processing and information theory. (Email: renxunhuan@bsuir.by)

Viktar Yurevich Tsviatkou received the Ph.D. degree from the Belarusian State University of Informatics and Radioelectronics, in 1999. He works at the Belarusian State University of Informatics and Radioelectronics, where he is currently a Professor and the Dean of the department of infocommunication technologies, Faculty of Information Security. His research interests include digital image processing, pattern recognition, signal processing, and information theory. (Email: vtsvet@bsuir.by)