PPDM-YOLO: A Lightweight Algorithm for SAR Ship Image Target Detection in Complex Environments

Hongjie He, Tianwen Hu, Sheng Xu, Hong Xu, Lin Song, and Zhongpei Sun

***Abstract*—To address the critical challenges in synthetic aperture radar (SAR) ship target detection, including complex background speckle noise interference and the difficulty in balancing model lightweight design with detection accuracy, this article proposes an innovative PPDM-YOLO model. Through modular architecture design, we establish a four-part technical framework: First, a lightweight feature extraction module named PCA is developed to reduce computational complexity by analyzing feature map redundancy, effectively mitigating feature degradation caused by noise. Second, the noise-resistant enhancement module, PSA-G, integrates the multiscale adaptive gradient threshold module with a dynamic spatial attention mechanism. This integration enhances target feature representation while effectively suppressing noise interference. Third, DySample technology is employed in place of conventional upsampling methods to improve the quality of feature reconstruction and preserve spatial details. In addition, a multiscale fusion small target detection network is introduced to boost small object detection through cross-layer feature interaction. Experimental results on HRSID and SSDD datasets demonstrate that PPDM-YOLO achieves 93.7% mAP50 and 70.3% mAP50–95 on HRSID, while reaching 99.4% mAP50 and 78.7% mAP50–95 on SSDD, showing significant advantages over mainstream detection models. With 34.7% fewer parameters than YOLOv11n, our model achieves optimal balance among noise suppression, model lightweighting, and detection accuracy. This research provides an efficient and reliable technical solution for real-time SAR ship detection in complex marine environments.**

***Index Terms*—Deep learning, synthetic aperture radar (SAR), ship detection, YOLOv11n.**

I. Introduction

SYNTHETIC aperture radar (SAR) plays a pivotal role in maritime surveillance, particularly in vessel detection, due to its all-weather and all-lighting imaging capabilities. Unlike optical sensors, SAR is independent of weather conditions, enabling the stable acquisition of high-resolution images. This allows for precise capture of vessel shapes and structural features, providing reliable support for marine safety and traffic monitoring. Through its multimode observation capabilities, SAR can efficiently achieve dynamic monitoring of vessels across different sea areas, making it a key technology for military reconnaissance and marine control [1], [2], [3], [4].

Conventional SAR ship detection approaches predominantly utilize the constant false alarm rate (CFAR) algorithm and its improved variants [5]. For example, Zeng et al. [6] combined CFAR processing with dual-polarization data to leverage the rich feature representation of dual-polarization and the intensity suppression of background noise by CFAR. This method improves the ability to detect small-scale maritime targets and emphasizes potential ones. Li et al. [7] proposed a two-step CFAR detection architecture that focuses on identifying target superpixels through a combination of global and local detection mechanisms. The statistical properties of superpixels were described using weighted information entropy, enabling better differentiation between target and clutter superpixels. CFAR detection operates by exploiting the scattering differences between targets and background. This method is effective for strong scattering targets when prior target information is unavailable. However, it faces limitations in adapting to complex sea conditions, with issues such as a high false-negative rate for diminutive targets and poor detection stability, which restrict its practical application.

In recent times, the swift progress in deep learning technologies has led to neural network-based detection methods gradually outperforming traditional CFAR algorithms. Deep learning detection technologies are typically divided into two primary categories: single-stage detectors [8] and two-stage detectors [9]. Among them, you only look once (YOLO) [10] and single shot multibox detector (SSD) [11] are the most prevalent single-stage detectors, while region-based convolutional neural networks (R-CNN) [12], fast R-CNN [13], and faster R-CNN [14] are typical representatives of two-stage detectors. Two-stage detectors offer higher accuracy but face significant bottlenecks in ship recognition speed. In contrast, single-stage detectors strike a better balance between detection speed and accuracy, delivering enhanced real-time performance.

SAR ship target detection across multiple scales remains a prominent research focus. Zhao et al. [15] incorporated the

Received 25 March 2025; revised 8 May 2025 and 20 July 2025; accepted 18 August 2025. Date of publication 25 August 2025; date of current version 15 September 2025. This work was supported in part by the Henan Provincial Science and Technology Research Project under Grant 252102210001, and in part by the Youth Fund of the National Natural Science Foundation of China under Grant 62201512. *(Corresponding author: Sheng Xu.)*

Hongjie He, Sheng Xu, and Hong Xu are with the School of Electronics and Information, Zhengzhou University of Light Industry, Zhengzhou 450002, China (e-mail: hhj@zzuli.edu.cn; xusheng@zzuli.edu.cn; xhspla@163.com).

Tianwen Hu, Lin Song, and Zhongpei Sun are with the School of Computer Science and Technology, Zhengzhou University of Light Industry, Zhengzhou 450002, China (e-mail: 332305040477@zzuli.edu.cn; songlin@zzuli.edu.cn; 332305050489@zzuli.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2025.3602497

convolutional block attention module (CBAM) focus mechanism into the feature pyramid network (FPN) to boost multiscale ship detection in complex SAR environments.Wang et al. [16] combined asymmetric pyramid nonlocal blocks and SimAM attention to mitigate nearshore background clutter. After the C3 module output, a channel transformation enhanced interchannel communication, boosting the detection of multiscale targets and small object accuracy. Zhu et al. [17] proposed DB-FPN, which improves the multiscale detection capability of ships by enhancing the integration of spatial and semantic data and by fully utilizing feature maps from different locations through feature reuse. Si et al. [18] put forward an optimized bidirectional feature fusion network architecture during the feature fusion stage. By integrating bidirectional cross-layer pathways (top-down and bottom-up), this framework dynamically fuses multiscale features through adaptive weighting, thereby achieving significant improvements in ship detection accuracy under multiscale scenarios. Liangjun et al. [19] introduced MSFA-YOLO, which integrates the DenseASPP module to enhance feature extraction for large ships with a larger receptive field, strengthening the model's adaptability to multiscale features. Huang et al. [20] developed alpha IOU, which introduces a weighted combination into the traditional IOU calculation, allowing for flexible adjustment of the loss function, enabling the model to better adapt to variations in object scale and shape. Wang et al. [21] proposed MSDNet, which effectively utilizes ship target information from low-level feature maps, allowing the detector to better concentrate on tiny vessels. The fused multiscale features retain abundant contextual details. Hong et al. [22] combined visible light images with SAR ship images for multiscale detection, boosting the model's robustness in detecting multiscale targets. Zhang et al. [23] combined deep networks with internal scattering features of ship targets to make better use of scattering information, thereby improving the ability to represent global features. Gong et al. [24] introduced the SSPNet network, based on FPN and faster R-CNN, which achieved excellent results in extracting multiscale features. Liu et al. [25] proposed DSMF-Net, which enhances the feature fusion stage by integrating a selective feature fusion (SFF) module with the MSCA attention mechanism. This design enables the network to flexibly align high-level semantic features with low-level spatial features, thereby improving detection performance under complex backgrounds and for multiscale ship targets in SAR imagery. Sun et al. [26] proposed two feature balancing modules: the MSLK-Block and the DFF-Block. The MSLK-Block integrates large-kernel convolutions with partitioned heterogeneous operations to enable efficient extraction of multiscale features. The DFF-Block employs a dynamic adaptive mechanism to fuse features across spatial and channel dimensions, effectively enhancing feature representation and discrimination. Together, these modules significantly improve the detection performance of ship targets in SAR imagery.

As the demand for real-time detection in practical applications continues to grow, lightweight detection models have emerged as a significant research focus in recent years. Feng et al. [27] put forward NLCNet, which uses the depthwise separable convolution (DSC) introduced in MobileNetV1 as the basic building block, reducing the model size and improving detection speed. Mao et al. [28] put forward the hierarchical feature fusion and attention network (HFFANet), which achieves fast, automatic, and high-accuracy ship target detection results through more efficient feature extraction and effective multilevel feature fusion. Xiong et al. [29] enhanced the C3 backbone of YOLOv5n by embedding squeeze-and-excitation (SE) and CBAM attention mechanisms, resulting in the C3SE and C3CBAM modules. These units adaptively recalibrate channelwise and spatial feature responses, achieving a balance between speed of computation and accelerated inference performance in real-time scenarios. Guo et al. [30] put forward the DBA module, which integrates DSC, batch normalization (BN) layers, and ACON activation functions. Through lightweight architectural design and dynamic parameter optimization strategies, this module effectively reduces model complexity while significantly accelerating convergence during training. Yin et al. [31] optimized the pointwise convolution component in DSC by replacing it with grouped pointwise convolution. In addition, integrating a channel attention mechanism into the convolutional unit design enables the method to achieve model lightweighting while marginally improving object detection accuracy. Cui et al. [32] suggested a lightweight TNN network with an optimal segmentation approach, using the harmonic mean of precision and recall to segment ship targets. This model enables rapid ship detection in large-scale SAR images. Li et al. [33] reorganized the feature extraction network in faster-RCNN by employing a preliminary architecture, thereby decreasing network complexity and parameter quantity while amplifying the effect of multiscale feature fusion. Hao and Zhang [34] redesigned a lightweight network module, MobileNetV3S. Combining MobileNetV3S with the cross stage partial network created a lightweight backbone network, achieving a slight accuracy improvement while significantly reducing the number of parameters. Xu et al. [35] introduced a lightweight cross-stage partial module, which reduces computational complexity and optimizes model compactness through integrated network pruning techniques. Tang et al. [36] proposed BESW-YOLO, which integrates a lightweight EMSC-C2f module into the feature extraction network. This module enhances feature extraction for targets of varying scales through multiscale convolution, while significantly reducing computational overhead. As a result, it improves both the efficiency and effectiveness of multiscale ship detection in SAR imagery. Man and Yu [37] proposed the NSFE module based on SPDConv and MobileNetV3 to enhance feature extraction capability. In addition, they designed the MECA attention mechanism, which integrates both global and local information. This design effectively strengthens multiscale perception and maintains detection accuracy, while significantly reducing the number of parameters and computational cost. Hao et al. [38] proposed a lightweight backbone network, CSP-MobileNetV3_UP, which significantly reduces the number of parameters and computational complexity while maintaining strong feature extraction capabilities. In addition, the designed MSFE-SAR module combines dilated convolution, attention mechanisms, and a feature pyramid structure to expand the receptive field and enhance multiscale feature fusion with minimal computational overhead, enabling efficient detection of multiscale targets in SAR imagery.

In addition, arising from the unique imaging mechanism of SAR, speckle noise is inevitably generated, which significantly impacts targets, particularly nearshore ship targets and those in close proximity to one another. This results in more severe false negatives and false alarms. To address this issue, The N-YOLO framework developed by Tang et al. [39] incorporated a SAR target potential area extraction module, which synergistically combines the CA-CFAR algorithm with morphological dilation operations to ensure the comprehensive extraction of suspected target regions in SAR imagery. The detection module optimizes subsequent detection processes by fusing the original image with extracted potential regions, generating a low-noise reconstructed image for enhanced analysis. Zhao et al. [40] proposed a simplified morphological denoising module, Sim-Mor, which effectively suppresses speckle noise with minimal computation, enabling the module to be integrated into the backbone. This allows the network to concentrate more on the features of the target. Zhao et al. [41] also introduced an imaging preprocessing structure that combines deep morphological networks, providing the detection network with features that contain edge information and reduced noise. Dai et al. [42] proposed DenoDet, a network that focuses more on high-frequency transformation calibration convolutional biases. It forms a natural multiscale subspace representation for target detection from the perspective of denoising across multiple subspaces.

Despite significant advancements in multiscale detection techniques for SAR ship target recognition, several critical challenges remain in optimizing detection performance.

1) *Target feature loss:* Inherent speckle noise in SAR images causes the model to lose critical ship features, such as edge sharpness, scatter point distribution, and structural information, during training, leading to degraded feature representation capability.
2) *Insufficient feature extraction:* The low signal-to-noise ratio environment hinders CNNs from capturing complete features—speckle noise obscures target textures, small targets suffer from feature degradation due to network downsampling, and background noise introduces feature interference.
3) *Efficiency-accuracy trade-off:* While existing lightweight models reduce parameter complexity, they often exacerbate shallow feature loss, further compromising detection performance.

These challenges are interrelated: the first two directly limit detection accuracy, while the third restricts practical deployment. Therefore, developing a noise-robust feature extraction framework that balances accuracy and efficiency remains a critical research frontier in SAR target detection.

To address the aforementioned technical challenges, this study innovatively proposes the PPDM-YOLO lightweight detection architecture. Through synergistic optimization of multiple modules, it achieves an optimal balance between computational efficiency and detection accuracy. Extensive experimental validation on representative SAR ship datasets such as HRSID [43] and SSDD [44] has demonstrated its superior performance. The main technical contributions of this work are manifested in the following aspects:

1) To effectively mitigate feature loss caused by speckle noise in SAR images and reduce the computational burden introduced by redundant information, we propose a redundancy-aware PCA module. By integrating partial convolution (PConv) with a channel attention mechanism, this module enhances noise robustness while compressing redundant features, enabling lightweight and efficient feature extraction.
2) To balance noise robustness and feature representation in SAR images, this article introduces the PSA-G module. It constructs a 3-D gradient space based on multiscale adaptive gradient threshold (AGT), employs a lightweight threshold generator alongside normalized soft thresholds to suppress noise, and integrates a dynamic point-state attention block (DPSA) to model feature dependencies. Through gradient-domain noise suppression and dynamic calibration mechanisms, the PSA-G module significantly enhances the accuracy and anti-interference capabilities of SAR small target detection in complex sea conditions.
3) We replace the original YOLOv11n upsampling layer with the DySample module in the neck, enabling flexible point-based interpolation. This enhances multiscale feature reconstruction and better preserves spatial details for small target localization.
4) We propose multiscale fusion small target detection network (MSTFNet) module, a lightweight multiscale fusion framework that integrates high-level semantics and low-level detail features. It improves detection performance for dense and small ships while keeping computational cost low—suitable for real-time SAR scenarios.

II. Related Work

A. YOLOv11n

The YOLOv11n [45] architecture consists of three core components: a backbone network, a neck network, and a head network. The backbone employs deep convolutional operations to extract hierarchical features from input images, while the neck optimizes cross-layer feature fusion by adaptively integrating multiscale features from the backbone, thereby enhancing the efficiency of semantic information propagation. The head performs the tasks of object classification and localization. As shown in Fig. 1, YOLOv11n introduces a novel architectural design in these three parts, with the most notable innovations being the C3K2 convolution mechanism and the C2PSA module. The C3K2 module is integrated into multiple channels within the head to handle multiscale features across different depths. Its structure adapts based on the C3K parameter: when C3K = False, it functions similarly to the C2f module, employing a standard bottleneck design. Conversely, when C3K = True, the bottleneck structure is substituted with the C3 module, facilitating deeper and more complex feature extraction. Key attributes of the C3K2 block include improved efficiency—by replacing a single large convolution with two smaller ones, computational overhead is minimized, accelerating feature extraction. In addition, YOLOv11n introduces architectural improvements to the C2 module by integrating pointwise spatial attention

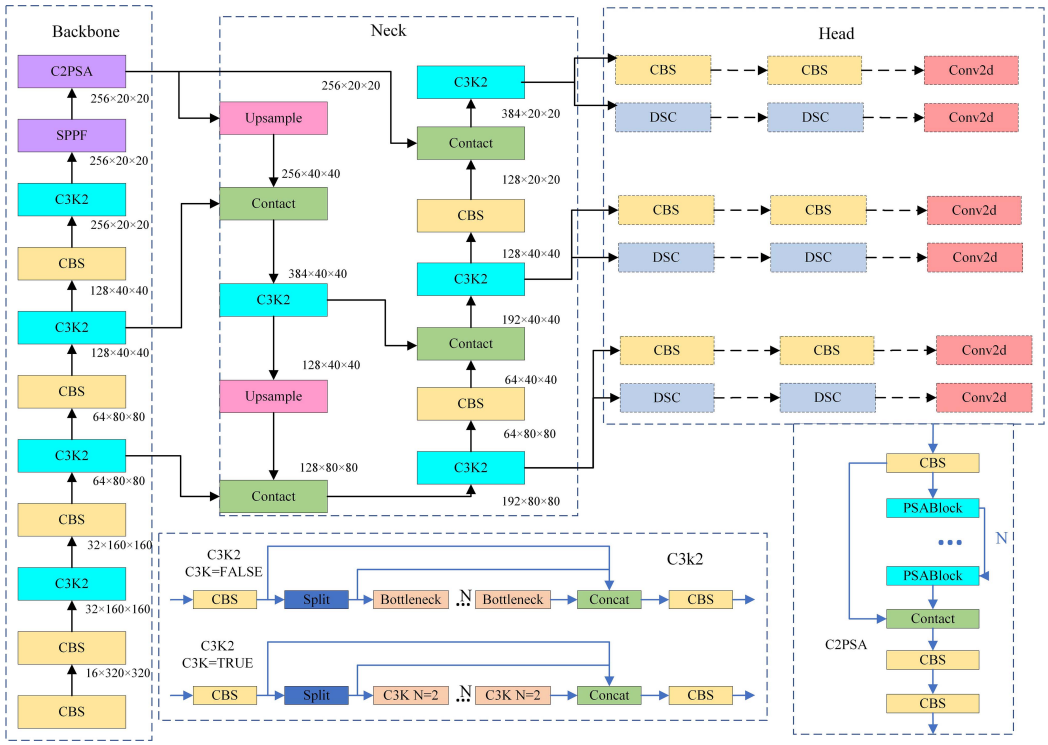

Fig. 1. YOLOv11n algorithm framework.

(PSA) to construct the C2PSA module, significantly enhancing core feature extraction capabilities. Building upon the multihead attention mechanism design strategy, this module demonstrates notable improvements in small-scale object recognition and detection accuracy in complex scenarios. The decoupled detection head at the network's terminal has been structurally optimized through the incorporation of dual DSC module. This enhancement effectively reduces parameter size and computational complexity while maintaining model compactness, achieving an optimal balance between detection efficiency and computational resource utilization.

B. Partial Convolution

PConv is an improved convolution operation designed for image inpainting and missing data handling. Its core idea is to adaptively focus on the valid pixel regions through a dynamic mask update mechanism [46]. Let the input feature map be $X \in \mathbb{R}^{C\times H\times W}$, with the corresponding binary mask $M \in \{0,1\}^{H\times W}$ (where 1 indicates valid pixels and 0 indicates missing pixels), the convolution kernel weights be $W \in \mathbb{R}^{k\times k\times C\times C'}$, and the bias be $b \in \mathbb{R}^{C'}$. The output Y and the updated mask M' of the PConv are calculated as follows:

$$Y(p) = \begin{cases} \dfrac{\sum_{ij} W_{ij}\cdot X(p+i,p+j)\cdot M(p+i,p+j)}{\text{sum}(M_{\text{patch}})} + b, & \text{if sum}(M_{\text{patch}}) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

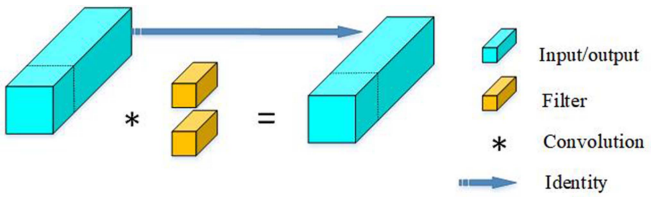

Fig. 2. PConv.

where p represents the output position coordinates, and denotes the mask region within the current convolution window. The normalization factor $\text{sum}(M_{\text{patch}})$ ensures that the output value is not affected by the number of valid pixels. After each convolution, the mask is updated as follows:

$$M'(p) = \begin{cases} 1, & \text{if sum} > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The updated mask is passed to the next layer, gradually reducing the influence of noise regions. Specifically, PConv selectively applies standard convolution operations to a subset of input channels for spatial feature extraction, while simultaneously maintaining the original feature representations of the unprocessed channels. This design significantly reduces computational complexity. Moreover, by reducing memory access, PConv is more suitable for I/O-constrained devices. Fig. 2 describes the working principle of PConv: it applies the filter only to a few input channels while keeping the other channels unchanged, thereby achieving efficient computation.

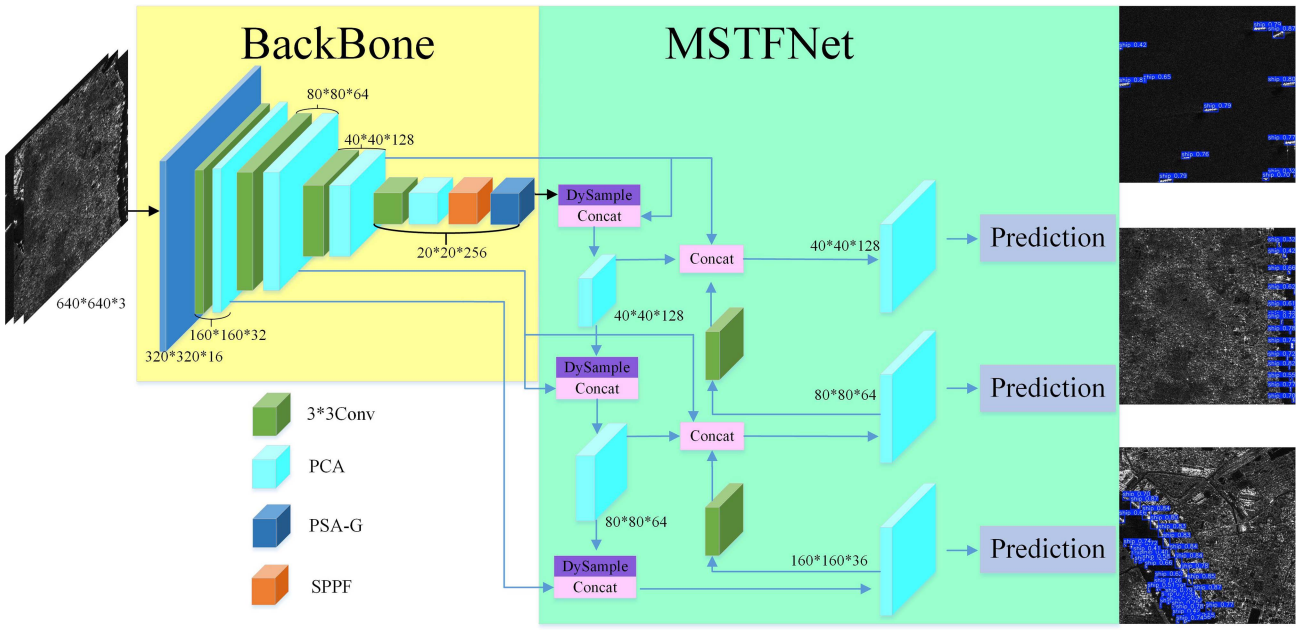

Fig. 3. PPDM-YOLO network structure.

III. Proposed Method

A. Overview of the Algorithm

This article introduces a more efficient SAR ship detection network framework, PPDM-YOLO, based on YOLOv11n, as shown in Fig. 3. The input is a 640 × 640 SAR ship image. In the backbone network feature extraction, we propose a more efficient PCA module that selectively strengthens the effective features of ship targets while ignoring noise and irrelevant information to reduce channel redundancy, thereby reducing both parameters and computational complexity. In addition, to further improve the feature extraction capabilities of PCA, we adopt the efficient channel attention (ECA) network [47], which provides an optimized balance between model complexity and channelwise feature calibration capability. Building on this, we integrate the PSA-G module for effective noise suppression and efficient feature fusion, enhancing the overall feature representation. The introduction of DySample [48] focuses more on the fine details of small targets in SAR ship images, avoiding the potential target information loss and inefficient feature learning typically found in traditional upsampling structures. The proposed methodology significantly enhances the identification accuracy of small-scale maritime targets amidst cluttered marine environments while mitigating interference from heterogeneous background elements. MSTFNet, combined with the PCA and DySample modules, for enhanced precision in detecting smaller ship targets, with the final detection processed by the head network.

B. PCA Module

Ship targets in large-scale scenes are often interfered with by sea surface clutter, lighting pollution, and other disturbances, especially when docked near coastal piers where the ship's outline is similar to that of the dock structures. The subtle feature differences can easily be lost during information transmission. Traditional convolution operations compute pixel values over the entire convolution kernel, but when some regions of the image have missing data (such as occlusions or noise), traditional convolutions may produce inaccurate results. To overcome this issue, we propose replacing traditional convolution with PConv and designing a lightweight feature extraction module, PCA, that integrates the ECA mechanism, as shown in Fig. 4. In the PCA module, the in.ut feature map is originally handled for preliminary feature extraction, followed by halving the number of channels to effectively reduce computational load. The remaining spatial features are extracted using PConv operations. The ECA attention mechanism is applied at the front end of the PConv to enhance the feature representation. Finally, the features extracted by PConv are concatenated with the retained features along the channel dimension and fused using a pointwise convolution, further reducing the computational load and generating the final output. The PCA module enhances feature extraction capability for ship targets while ensuring efficiency, particularly in scenarios with noise or missing data.

C. PSA-G Module

To address the challenges posed by speckle noise interference, multiscale feature representation difficulties of ship targets, and high false alarm rates caused by complex ocean backgrounds in SAR images, this article proposes a PSA-G module based on adaptive gradient thresholding and dynamic attention. The proposed framework establishes an optimal equilibrium between noise robustness and feature discrimination in SAR ship detection through gradient-domain noise suppression, feature decoupling learning, and dynamic calibration mechanisms. As shown in Fig. 5, at the network architecture level, we introduce

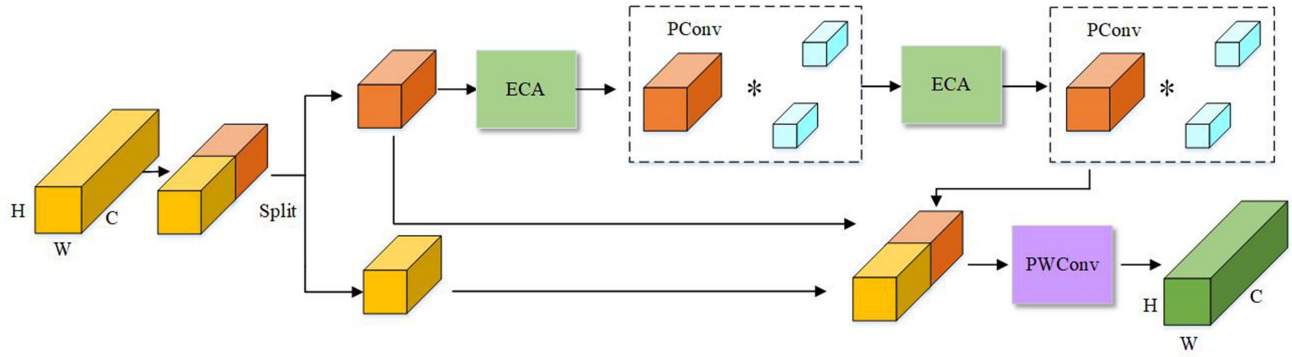

Fig. 4. PCA Module.

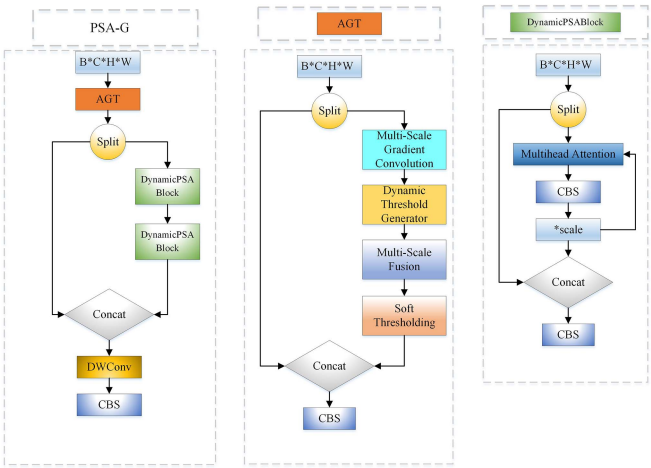

Fig. 5. PSA-G Module.

a multiscale AGT module. This module constructs a 3-D feature space using fixed-parameter Sobel operators for horizontal, vertical, and diagonal gradients. A lightweight threshold generator dynamically predicts spatially adaptive thresholds, and a max-normalized soft-thresholding function is employed to truncate noise gradients. Given an input feature map $X \in \mathbb{R}^{C\times H\times W}$, the AGT module constructs a 3-D gradient space using a fixed-parameter Sobel operator set $\{\nabla_k\}_{k=1}^{3}$.

$$\begin{cases} \nabla_1 = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} & \text{(horizontal)} \\ \nabla_2 = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} & \text{(vertical)} \\ \nabla_3 = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix} & \text{(diagonal).} \end{cases} \tag{3}$$

The multiscale gradient features are as follows:

$$G = [\nabla_1 X, \nabla_2 X, \nabla_3 X] \in \mathbb{R}^{3C\times H\times W}. \tag{4}$$

G is processed by a lightweight threshold generator to predict spatially adaptive thresholds:

$$\tau = \sigma(\mathcal{W}_2(\text{ReLU}(\mathcal{W}_1(G)))) \tag{5}$$

where $\mathcal{W}_1 \in \mathbb{R}^{3C\times C/4}$ and $\mathcal{W}_2 \in \mathbb{R}^{C/4\times 3C}$ are 1×1 convolution layers, and σ represents the Sigmoid function. Finally, noise suppression is achieved through a max-normalized soft-thresholding function as follows:

$$\mathcal{S}(g_{ijk}) = \text{sign}(g_{ijk}) \cdot \text{ReLU}\left(|g_{ijk}| - \tau_{ijk} \cdot \max_c(|g_{ijc}|)\right) \tag{6}$$

Next, during the feature processing stage, we introduce a DPSA Block. After the input feature map undergoes multihead self-attention processing, the generated attention output is scaled by a learnable factor α. This scaling factor is optimized through backpropagation, allowing the model to automatically adjust the influence of the attention mechanism during training. This adaptability enhances the model's ability to accommodate the diversity and complexity of targets in SAR images. The processing of an input feature map $B \in \mathbb{R}^{C'\times H\times W}$ is able to be formulated as follows:

$$\text{DPSA}(B) = B + \alpha \cdot \text{MHA}(B). \tag{7}$$

During the feature fusion stage, a dual-stream decoupled learning strategy is employed, where the dimension-reduced feature space is orthogonally decomposed into a low-frequency contextual branch (Branch a) and a high-frequency detail branch (Branch b). Branch a utilizes DSC to preserve the overall structure of the target, while Branch b enhances local saliency features through a DPSA sequence. Finally, cross-scale feature interaction is achieved via grouped convolution. This module effectively addresses the diversity and challenges of SAR ship images, demonstrating superior performance and enhanced stability, particularly in complex background detection tasks.

D. DySample Module

The upsampling method used in YOLOv11n, nearest neighbor interpolation (NNI), is commonly applied to enlarge the feature map dimensions. However, due to its fixed rule, it cannot adaptively adjust sampling, limiting its performance in complex tasks. Although dynamic upsampling methods, such as content-aware reassembly of features (CARAFE) [49], efficient

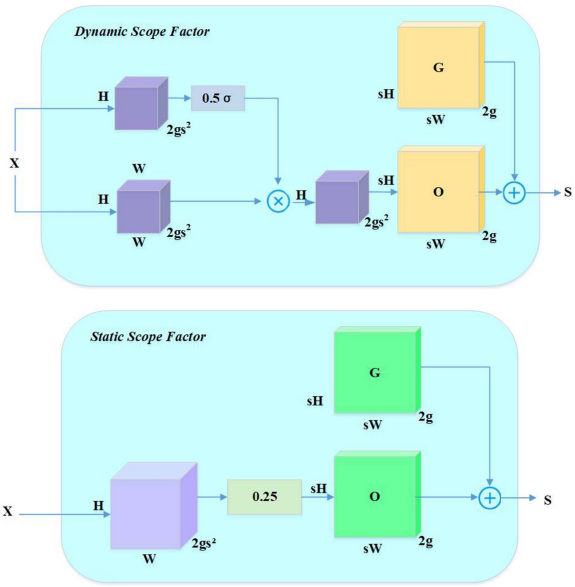

Fig. 6. DySample Module.

upconvolution block (ECUB) [50], and SAPA [51], can enhance performance, they often require complex dynamic convolutions or additional subnetworks, which increase computational and memory overhead, thus limiting practical applications.

To improve efficiency and reduce computational resource consumption, we introduce the lightweight DySample module in the feature extraction of the neck network. Unlike traditional dynamic upsampling methods, DySample redefines the upsampling process from a point-based sampling perspective, avoiding the complex computations that rely on convolutional kernels. By generating offset positions to sample feature points, it enhances the input features and effectively reduces computational complexity.

The core concept of DySample is to return to the essence of upsampling—point sampling—to simulate geometric information in the feature map. For an upsampling scale factor s and a feature map X of size $C \times H \times W$, a linear layer transforms the input with C channels into an output with $2\,s^2$ channels, generating an offset O of size $2\,s$. This offset is then reshaped into $2 \times sH \times sW$ through pixel transformation. To increase the flexibility of the offset, a linear projection of the input features is used to generate pointwise "dynamic range factors." By applying the sigmoid function and a static factor of 0.5, the dynamic scope maps values within the range [0, 0.5] to a static value centered at 0.25. As shown in Fig. 6, the bottom box illustrates the version with the "static range factor," where the offset is generated through a linear layer. The top box presents the "dynamic range factor" version, in which the range factor is first derived and then used to adjust the offset. The symbol σ denotes the sigmoid function. The sampling set S is obtained by adding the offset O to the original sampling grid G, approach as follows:

$$X = \text{grid\_sample}(X, S) \tag{8}$$

$$S = O + G \tag{9}$$

$$O = 0.5 \cdot \sigma(\text{liner1}(X) \cdot \text{liner2}(X)). \tag{10}$$

The introduction of DySample places greater emphasis on the fine details of small targets in SAR ship images, addressing the issues of target information loss and inefficient feature learning that often occur with traditional upsampling structures. This results in improved performance for small ship target detection in complex backgrounds.

E. MSTFNet Module

In our approach, we propose a MSTFNet that considers features from both the same level and adjacent levels. This approach allows high-level information to be transferred to lower level features, thereby retaining rich semantic information. We introduced an ultrasmall 160 × 160 target detection layer in the head network to address the loss of small target information in high-level features. In addition, considering that large ship targets are less common in SAR images, we made adjustments for large target detection by removing the original large target detection head for the 20 × 20 feature map and replacing it with a mid-sized target detection head for the 40 × 40 feature map. This change allows for the detection of both large and medium targets while maintaining stable accuracy and reducing the number of parameters.

As shown in Fig. 7, we introduce a lateral fusion strategy among multiscale feature maps to enhance the detection capability for targets of varying scales. Specifically, the medium-scale detection head with a resolution of 40 × 40 is enhanced by concatenating features from PCA3, PCA5, and Conv6, thereby improving the recognition of medium and large ships. The small-scale detection head with a resolution of 80 × 80 integrates features from PCA2, PCA6, and Conv5 to improve the perception of small ships. In addition, by concatenating DySample3 with PCA1, we introduce a new ultrasmall object detection head to strengthen the model's performance in detecting high-resolution, fine-grained small targets. Ultimately, three detection heads are formed with the following specific functions:

PCA7: A newly added ultrasmall object detection head designed for 160 × 160 high-resolution feature maps, specifically aimed at improving the detection accuracy of ultrasmall ships in SAR images.

PCA8: A small object detection head based on 80 × 80 resolution feature maps, providing enhanced capability for detecting small ships.

PCA9: A medium-to-large scale object detection head operating on 40 × 40 resolution feature maps, primarily used for detecting common medium and large ships.

This method optimizes cross-scale feature connectivity and enhances the model's ability to effectively capture and utilize hierarchical information, while also extracting shallower convolutional features to capture the fine details of small ship targets.

$$\text{PCA7} = \text{concat}(\text{DySample3}, \text{PCA1}) \tag{11}$$

$$\text{PCA8} = \text{concat}(\text{PCA2}, \text{PCA6}, \text{Conv5}) \tag{12}$$

$$\text{PCA9} = \text{concat}(\text{PCA3}, \text{PCA5}, \text{Conv6}) \tag{13}$$

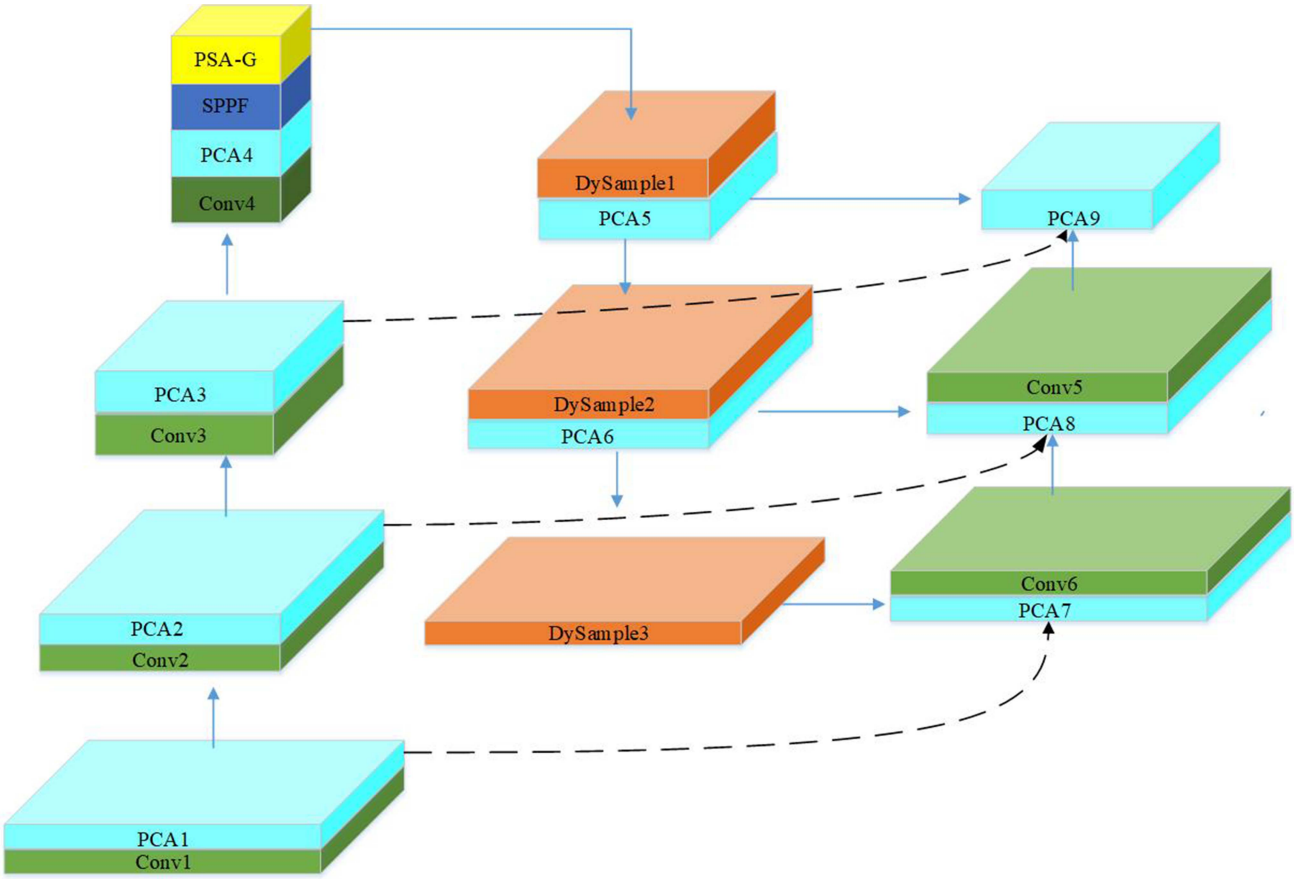

Fig. 7. MSTFNet Module.

IV. Experimental Results and the Analysis

A. Experimental Detail

All experiments were conducted on the same computer configuration, utilizing Python 3.9 and the PyTorch 2.1.2 framework. The hardware setup includes an Intel Core i7-12700F CPU @ 2.10 GHz and an Nvidia GeForce RTX 4060 Ti GPU. The experimental results of YOLOv11n were used as the baseline. During the training phase, the initial learning rate was set to 0.01, momentum was set to 0.9, and the AdamW optimizer (Adam with weight decay) was chosen to optimize the parameters of the neural network, enhancing stability and incorporating weight decay to reduce overfitting. The batch size was set to 16, and k-means clustering was employed to obtain multiscale anchor points matching the specific characteristics of the different datasets. The training ran for 300 epochs. During the testing phase, nonmaximum suppression (NMS) and confidence thresholds were configured at 0.45 and 0.25, respectively.

To enhance the scientific rigor and stability of model selection, our methodology incorporates an early stopping mechanism while maintaining a fixed upper limit of 300 training epochs to mitigate overfitting risks. The training automatically terminates when neither the mAP50 nor mAP50–95 validation metrics show improvement for 30 consecutive epochs. The model weights achieving peak validation performance (measured by both mAP50 and mAP50–95) are retained for final evaluation.

B. SAR Ship Dataset

The HRSID and SSDD datasets are two major SAR image datasets for remote sensing ship detection. The HRSID dataset, released by Wei et al. [43] in 2020, consists of 5604 high-resolution images (800 × 800 pixels, 0.5–3 m resolution) collected from TerraSAR-X, Sentinel-1, and TanDEM-X satellites. The dataset contains 16 951 annotated ship instances, with small (54.5%), medium (43.5%), and large (2%) ship bounding boxes covering 0%–0.16%, 0.16%–1.5%, and more than 1.5% of the image area, respectively, making it particularly challenging for small object detection. The SSDD dataset contains 1160 SAR images (500 × 500 pixels, resolution 1–15 m), integrating data from Sentinel-1, TerraSAR-X, and RadarSat-2 satellites, with a focus on small object detection in complex marine environments. Both datasets, with multisource data fusion and fine annotations, provide a multiscale, multiscenario benchmark platform for SAR-based ship detection algorithms.

In this experiment, the input image sizes for HRSID were set to 640 × 640, and for SSDD, they were set to 500 × 350. The data were randomly allocated to training and test groups in an

TABLE I
CONFUSION MATRIX

Truth	Forecast results	
	Positive example	Counter example
Positive example	TP	FN
Counterexample	FP	TN

8:2 proportion, and the model's performance was evaluated on small object detection and multiscale SAR ship detection tasks.

C. Evaluation Metrics

To analyze the results of object detection training, the model is evaluated based on the following metrics: mean average precision (mAP), Precision (P), Recall (R), the number of parameters (Params), frames per second (FPS), and the confusion matrix. The confusion matrix summarizes the true and predicted labels in a matrix form, as shown in Table I. The elements of the confusion matrix are defined as follows.

TP (True Positive): A correct identification of a positive instance by the model;

FN (False Negative): The model misses a true positive instance;

FP (False Positive): The model incorrectly flags a negative instance as positive;

TN (True Negative): A correct identification of a negative instance by the model.

1) Precision: The proportion of samples predicted as positive by the detection model that are actually positive. The formula for calculation is as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{14}$$

2) Recall: The proportion of actual positive samples that are correctly detected as positive. The formula for calculation is as follows:

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{15}$$

3) mAP: The average AP (average precision) value across multiple categories. The AP for each category is calculated first, and then the average value is taken as the overall performance metric. The formula for calculation is as follows:

$$\mathrm{mAP} = \frac{1}{c}\sum_{i=1}^{c} \mathrm{AP}_i \tag{16}$$

where c is the number of categories.

D. Comparison Experiments

To evaluate the strong detection capability of PPDM-YOLO for SAR images, we performed a comparative analysis under the same conditions and parameter settings with eight widely used SAR ship detection models: SSD [11], faster R-CNN [14], YOLOv5n [52], YOLOv8n [53], YOLOv11n [45], LMSD-YOLO [30], MSFA-YOLO [19], and BESW-YOLO [36]. Two datasets, HRSID and SSDD, were used for performance evaluation. The comparison results are shown in Tables II and III.

We conducted initial experiments using the HRSID dataset. Table II provides a detailed comparison of the accuracy for each model. Clearly, our model achieved the best performance, with an mAP50 of 93.7%. PPDM-YOLO improved the accuracy by 3.1% compared to YOLOv11n. We visualized near-shore targets in the HRSID dataset to demonstrate PPDM-YOLO's precise detection of these targets. As shown in Fig. 8, we selected three images with dense objects in complex scenes, which presented significant detection challenges. Despite these complexities, PPDM-YOLO consistently demonstrated its superior ability by accurately detecting objects in these challenging scenarios. The specific detection results for ship targets under different methods are shown in Table IV.

In the visualization results on the HRSID dataset, significant differences in the detection performance of the three baseline models were observed in complex near-shore backgrounds. In label Image 1, which contains 13 ship targets, the experiment shows that YOLOv5n produces four FP detections, while YOLOv8n and YOLOv11n both control the number of FP detections to three. In label Image 2, with nine targets, YOLOv5n results in 1 FP detection, but exhibits target splitting, where a single true target is misidentified as two separate targets. YOLOv8n produces two FP detections, and YOLOv11n optimizes the performance to only one FP detection. In the densely packed small target scene of label Image 3, which contains 11 targets, both YOLOv5n and YOLOv11n show one FN detection and one FP detection, while YOLOv8n maintains a relative advantage with only 1 FP detection.

Next, we conducted comparative experiments on the SSDD dataset. Compared to the HRSID dataset, the background in the SSDD dataset is relatively simpler. As shown in Table III, each model demonstrated high accuracy. Our model achieved a slight advantage with a mAP50 of 99.4%. In addition, the model's parameter count is 1.71 M, and its size is only 4.2 MB. Despite the minor accuracy differences, our model is the lightest while also achieving a moderate FPS. Similarly, as shown in Fig. 9, we visualized the near-shore targets in the SSDD dataset. Table V presents the detailed detection results of ship targets on the SSDD dataset under different methods.

In the visualization analysis of the SSDD dataset, notable differences were observed in the target detection performance of the seven models under complex near-shore backgrounds. In Label Image 1, which contains six ship targets, YOLOv5n exhibited the highest number of missed detections, producing four FN detections. YOLOv8n showed slightly better performance with two FN detections, while YOLOv11n further improved by reducing the FN detection to one but introduced two FP detections. In contrast, our model demonstrated the best performance, detecting nearly all targets with only one FN detection. In label Image 2, which consists of 11 ship targets, a similar trend was observed. YOLOv5n performed the worst, generating six FN detections, while YOLOv8n reduced one FN detection. YOLOv11n performed relatively well, producing only one FP. In label Image 3, which contains three ship targets, all three baseline models, YOLOv5n, YOLOv8n, and YOLOv11n, failed to achieve perfect detection, each producing one FP detection.

TABLE II
COMPARISON OF TARGET DETECTION WITH OTHER METHODS ON HRSID

Model	Dataset	P(%)	R(%)	mAP50(%)	mAP50-95(%)	Params(M)	FLOPs(G)	FPS	Size(MB)
SSD [11]	HRSID	92.3	58.7	79.9	64.0	–	–	21	–
Faster R-CNN [14]	HRSID	87.2	89.1	89.1	56.1	–	–	9	–
YOLOv5n [52]	HRSID	90.7	82.9	90.8	66.7	2.18	**5.8**	**333**	5.5
YOLOv8n [53]	HRSID	90.8	85.1	91.7	67.4	3.01	8.2	322	6.3
YOLOv11n [45]	HRSID	90.9	82.3	90.6	66.7	2.62	6.3	294	5.5
MSFA-YOLO [19]	HRSID	92.4	86.0	92.7	67.1	–	–	–	–
LMSD-YOLO [30]	HRSID	92.7	**86.4**	**93.9**	–	3.50	6.6	68	7.6
BESW-YOLO [36]	HRSID	92.3	81.2	90.0	–	**1.70**	6.9	–	–
Ours	HRSID	**93.4**	86.0	93.7	**70.3**	1.71	7.5	169	**4.2**

TABLE III
COMPARISON OF TARGET DETECTION WITH OTHER METHODS ON SSDD

Model	Dataset	P(%)	R(%)	mAP50(%)	mAP50-95(%)	Params(M)	FLOPs(G)	FPS	Size(MB)
SSD [11]	SSDD	93.2	72.8	94.4	57.0	–	–	21	–
Faster R-CNN [14]	SSDD	81.0	94.2	97.1	61.0	–	–	9	–
YOLOv5n [52]	SSDD	97.9	96.5	99.3	77.5	2.18	**5.8**	**333**	5.5
YOLOv8n [53]	SSDD	98.4	96.8	99.2	78.4	3.01	8.2	322	6.3
YOLOv11n [45]	SSDD	96.8	97.6	99.3	77.6	2.62	6.3	294	5.5
MSFA-YOLO [19]	SSDD	97.7	**98.0**	98.7	66.2	–	–	–	–
LMSD-YOLO [30]	SSDD	96.5	94.1	98.0	–	3.50	6.6	68	7.6
BESW-YOLO [36]	SSDD	95.2	92.2	97.3	69.6	**1.70**	6.9	–	–
Ours	SSDD	**98.4**	97.8	**99.4**	**78.7**	1.71	7.5	169	**4.2**

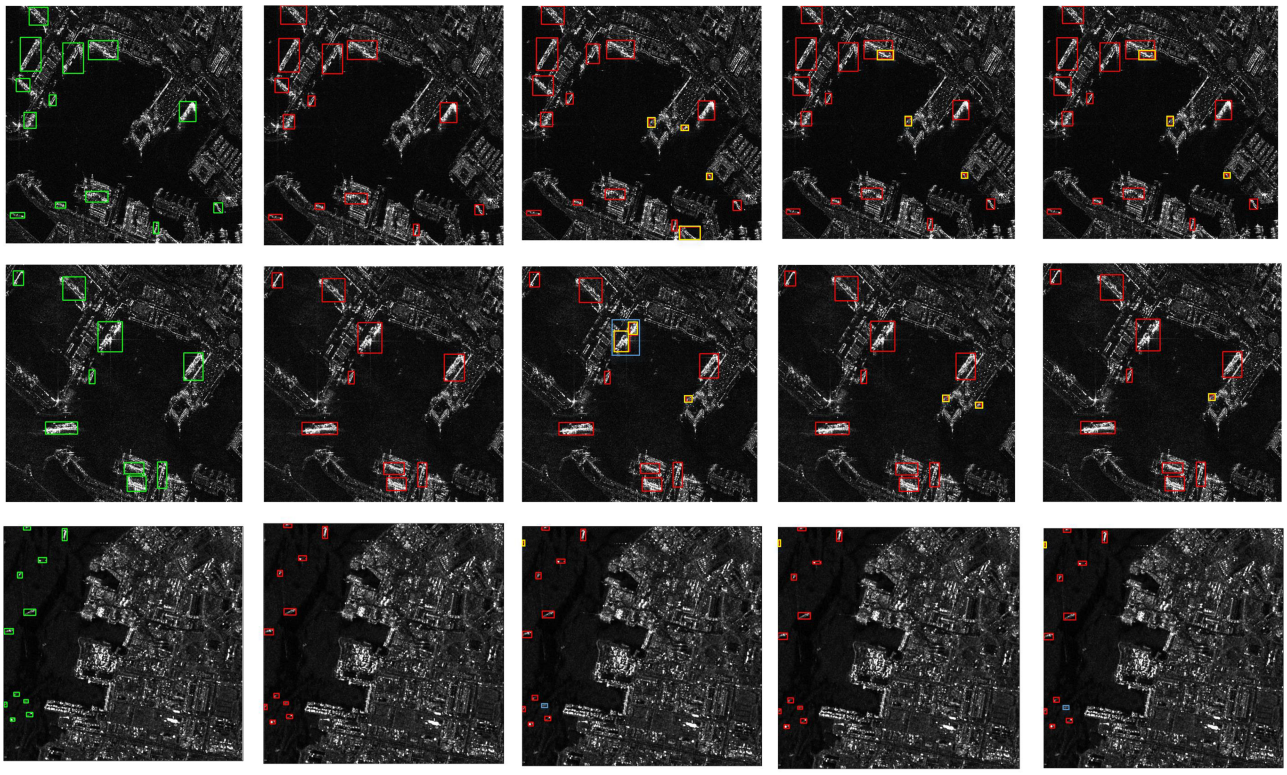

Fig. 8. Our method compares the visualization results with YOLOv5n and YOLOv8n in HRSID. (a) Ground truth (GT) image. (b) PPDM-YOLO. (c) YOLOv5n. (d) YOLOv8n. (e) YOLOv11n. Note that green boxes represent ground truth annotations, red boxes represent true positive ship targets, blue boxes represent FN ship targets, and yellow boxes represent FP ship targets.

TABLE IV
PERFORMANCE EVALUATION OF DIVERSE SHIP DETECTION APPROACHES ON HRSID

Method	Image1				Image2				Image3			
	GT	TP	FN	FP	GT	TP	FN	FP	GT	TP	FN	FP
YOLOv5n	13	13	0	4	9	8	1	3	11	10	1	1
YOLOv8n	13	13	0	3	9	9	0	2	11	11	0	1
YOLOv11n	13	13	0	3	9	9	0	1	11	10	1	1
Ours	13	13	0	0	9	9	0	0	11	11	0	0

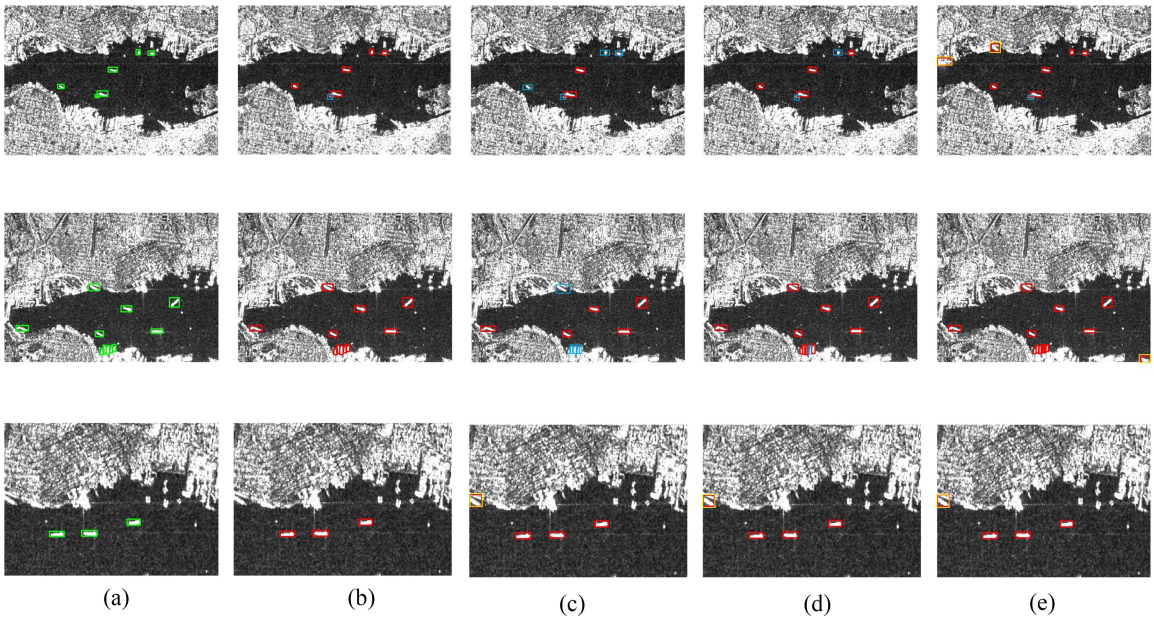

Fig. 9. Our method compares the visualization results with YOLOv5n and YOLOv8n in SSDD. (a) GT image. (b) PPDM-YOLO. (c) YOLOv5n. (d) YOLOv8n. (e) YOLOv11n. Note that green boxes represent GT annotations, red boxes represent TP ship targets, blue boxes represent FN ship targets, and yellow boxes represent FP ship targets.

TABLE V
PERFORMANCE EVALUATION OF DIVERSE SHIP DETECTION APPROACHES ON SSDD

Method	Image1				Image2				Image3			
	GT	TP	FN	FP	GT	TP	FN	FP	GT	TP	FN	FP
YOLOv5n	6	2	4	0	11	5	6	0	3	3	0	1
YOLOv8n	6	4	2	0	11	10	1	0	3	3	0	1
YOLOv11n	6	5	1	2	11	11	0	1	3	3	0	1
Ours	6	5	1	0	11	11	0	0	3	3	0	0

In our model, the PCA and PSA-G modules address issues such as target feature loss and insufficient feature extraction caused by speckle noise in SAR ship images with complex backgrounds. In addition, the introduction of MSTFNet and DySample significantly improves the detection accuracy of small targets. Furthermore, our model reduces the number of parameters, offering significantly better performance compared to other models.

E. Verification of Complex Background Dense SAR Images

To validate the performance of our method in dense target detection, we conducted comparative experiments using dense SAR ship images from two typical maritime scenarios: coastal and open-sea environments. Figs. 10 and 11 present the visual detection results for coastal and open sea SAR scenarios.

TABLE VI
COMPARISON RESULTS OF DIFFERENT MODELS FOR DENSE TARGETS IN HRSID UNDER OPEN SEA BACKGROUND

Method	GT	TP	FN	FP	P(%)	R(%)
YOLOv5n	51	44	8	1	97.8	84.6
YOLOv8n	51	48	7	4	92.3	87.3
YOLOv11n	51	45	9	3	93.8	83.3
Ours	51	49	2	0	100	96.1

Tables VI and VII systematically compare the detection metrics between our method and three typical CNN models. In open-sea large-scene multitarget detection (see Table VI),

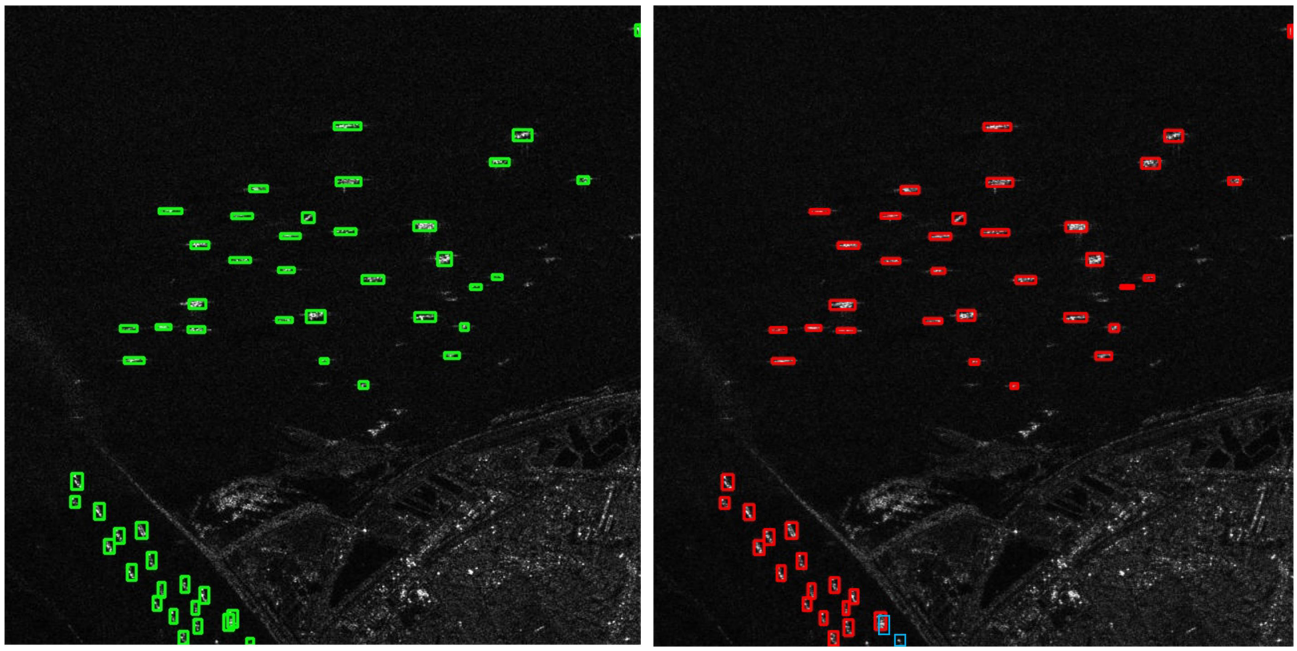

Fig. 10. Detection results of dense targets in SAR images of open sea scenarios. Note that green boxes represent GT annotations, red boxes represent TP ship targets, and blue boxes represent FN ship targets.

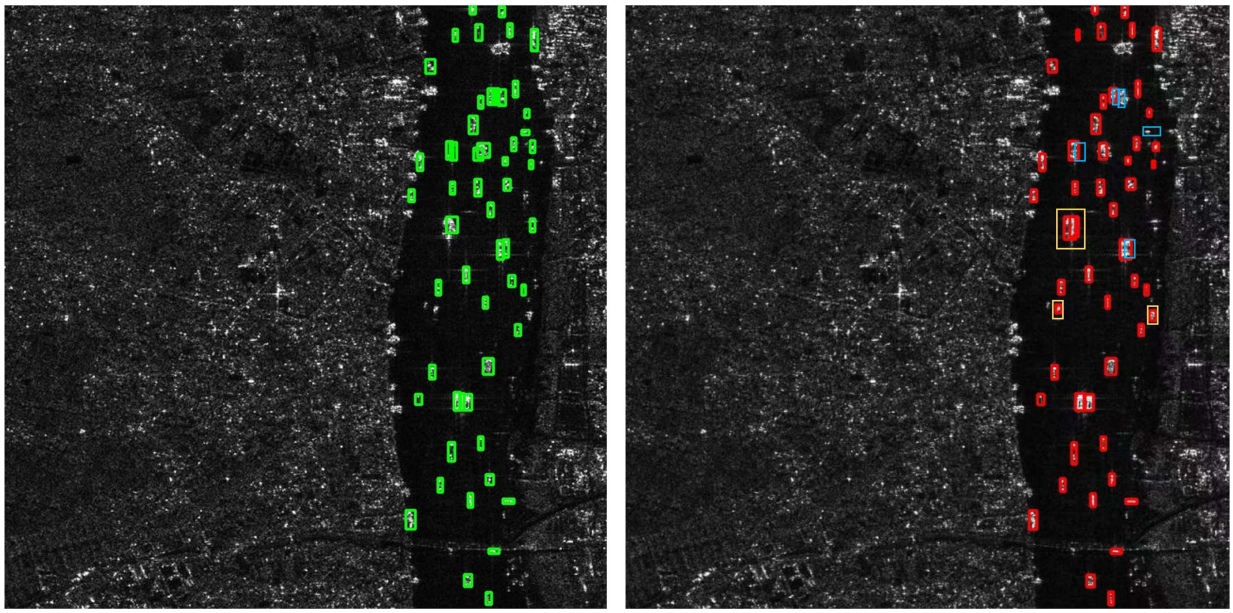

Fig. 11. Detection results of dense targets in SAR images for coastal scenarios.Note that green boxes represent GT annotations, red boxes represent TP ship targets, blue boxes represent FN ship targets, and yellow boxes represent FP ship targets.

our method successfully detected 49 ships with zero FP detections and only two FN detections, achieving 100% precision and 96.1% recall. This significantly outperforms comparison methods, demonstrating superior anti-interference capability in complex sea clutter environments. As shown in Table VII, our method maintains its leading position in more challenging high-resolution coastal SAR scenarios with 94.4% precision and 91.1% recall, benefiting from optimized feature extraction module and effective noise suppression strategy. However, three FN detections and five FP detections remain in coastal scenarios.

It should be emphasized that coastal detection presents significantly greater challenges than open-sea conditions. Quantitative analysis reveals performance degradation across all four models in coastal scenarios, attributable to two main factors: 1) Strong scattering interference from terrestrial structures and port facilities reduces target contrast; 2) Occlusion effects from dock

TABLE VII
COMPARISON RESULTS OF DIFFERENT MODELS FOR DENSE TARGETS IN HRSID UNDER OPEN SEA BACKGROUND

Method	GT	TP	FN	FP	P(%)	R(%)
YOLOv5n	53	40	16	3	93.0	71.4
YOLOv8n	53	39	20	6	86.7	66.1
YOLOv11n	53	41	15	3	93.2	73.2
Ours	53	51	5	3	94.4	91.1

TABLE VIII
PARAMETERS OF THE TWO MODELS ARE COMPARED ON THE BACKBONE

backbone	**YOLOv11n**		**PPDM-YOLO11**	
	Module	**Params**	**Module**	**Params**
1	Conv	464	Conv	464
2	Conv	4672	Conv	4672
3	C3K2	6640	PCA	4864 (-1772)
4	Conv	36 992	Conv	36,992
5	C3K2	26 080	PCA	18 496 (-7584)
6	Conv	147 712	Conv	147,712
7	C3K2	87 940	PCA	47 232 (-40 708)
8	Conv	295 424	Conv	295 424
9	C3K2	346 112	PCA	185 600 (-160 512)

TABLE IX
COMPARATIVE RESULTS OF DIFFERENT ATTENTION MECHANISM MODELS

Method	**P**(%)	**R**(%)	**mAP50**(%)	**mAP50-95**(%)
SE	91.4	82.1	91.5	66.9
CA	91.5	83.9	91.3	66.8
CBAM	90.7	82.7	90.7	67.2
PCA	92.7	86.0	92.5	68.4

installations or large vessels lead to incomplete target morphology, complicating feature identification. Our proposed algorithm addresses these challenges through multiscale feature fusion and context-aware modules, demonstrating stronger environmental adaptability than traditional CNN approaches.

F. Ablation Experiments

1) Effect of PCA Module: Furthermore, we investigated the lightweight design of the PCA module by comparing PPDM-YOLO based on the backbone with YOLOv11n. Table VIII shows that, in the first five layers, the advantage is not significant due to the lower number of channels. However, as the number of channels increases, our model demonstrates a substantial improvement in efficiency, particularly in the 7th and 9th layers, where it reduces 40 708 and 160 512 parameters, respectively. Overall, our model contains 1.71 M parameters, achieving a 34.7% reduction compared to YOLOv11n (2.62 M).

To further assess the effectiveness of the PCA model, we performed a comparative analysis with three alternative attention mechanism models: SE [54], coordinate attention (CA) [55], and CBAM [56]. The detailed experimental results are presented in Table IX.

TABLE X
COMPARATIVE RESULTS OF DIFFERENT UPSAMPLING MODULES

Method	**P**(%)	**R**(%)	**mAP50**(%)	**mAP50-95**(%)
NNI	90.9	82.3	90.7	66.7
EUCB	91.0	82.5	90.7	66.8
CARAFE	90.8	83.1	90.9	66.2
DySample	91.9	82.8	91.5	67.1

The data shows that, compared to the baseline model, mAP50 increased to 92.5%, and mAP50–95 increased to 68.4%, demonstrating the effectiveness of the lightweight network structure of the designed PCA module.

2) Effect of PSA-G Module: In Experiment 3, we investigated the role of the PSA-G module. Fig. 12 illustrates the changes in mAP50 and mAP50–95 values during training for both the baseline and our method. As shown by the blue curve for PSA-G, the mAP50–95 increased by 0.9%, and the mAP50 increased by 1%.

3) Effect of DySample Module: Table X presents the experimental results of different upsampling modules based on YOLOv11n. Compared with three mainstream upsampling approaches, CARAFE, EUCB, and NNI, our proposed method demonstrates significant improvements on the HRSID ship dataset. Specifically, it achieves a 1% increase in precision, 0.5% enhancement in recall, 0.8% gain in mAP50, and 0.4% improvement in mAP50–95 over the baseline model. The DySample module enables the model to concentrate on pivotal regions through dynamic feature adaptation, thereby boosting overall detection performance. Comparative analysis reveals that DySample not only outperforms existing upsampling techniques but also exhibits robust feature extraction capabilities and remarkable adaptability to diverse ship recognition scenarios, particularly in complex maritime environments with scale variations and occlusions.

4) Effect of MSTFNet Module: In Experiment 4, we investigated the role of the MSTFNet module. The data indicates that this module increased mAP50–95 by 3%, with mAP50 increased by 1.9%, while simultaneously reducing the parameter count by 0.66 M. The module employs lightweight feature fusion techniques to enhance efficiency and reduce computational load.

5) Summary: Finally, we combined all modules and conducted an ablation study on the HRSID dataset. As shown in Table XI, we used YOLOv11n as the baseline model to evaluate the effectiveness and impact of the PCA, PSA-G, DySample, and MSTFNet modules. The contribution of each module is marked with a "✓." Compared to YOLOv11n, our model reduced the parameter count by 34.7%, increased mAP50 by 3%, mAP50–95 by 3.6%, while maintaining a moderate FPS. These ablation studies highlight the significance of each module within the PPDM-YOLO framework, emphasizing their complementarity and their effectiveness in improving YOLOv11n performance.

V. DISCUSSION

With the growing demand for lightweight models and high-precision object detection, the proposed PPDM-YOLO in this

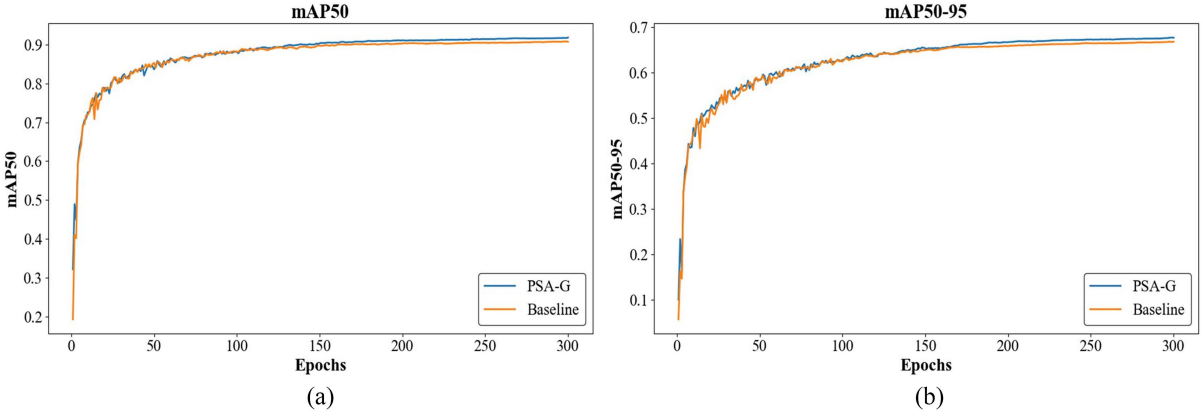

Fig. 12. (a) mAP50 curve. (b) mAP50–95 curve.

TABLE XI
ABLATION STUDIES OF MODEL COMPONENTS ON HRSID

Methods	DySample	PCA	PSA-G	MSTFNet	mAP50(%)	mAP50-95(%)	Params(M)	FPS
YOLOv11n					90.7	66.7	2.62	294
1	✓				91.5	67.1	2.63	312
2		✓			92.5	68.4	2.02	212
3			✓		91.7	67.6	2.60	263
4				✓	92.6	69.7	1.96	156
5	✓	✓	✓	✓	93.7	70.3	1.71	169

article provides an effective solution for SAR ship detection tasks, achieving a strong balance between accuracy and computational constraints. The model addresses common challenges in SAR images, such as speckle noise interference, significant variations in target sizes, and densely distributed small targets, by constructing a detection framework that features multimodule collaboration, strong feature representation, and suitability for resource-constrained environments.

First, the PCA module analyzes redundancy relationships among feature map channels and integrates the ECA attention mechanism to compress redundant features while enhancing critical information. This module not only effectively reduces model parameters and computational overhead but also improves the efficiency and robustness of feature extraction. Second, to tackle the severe noise interference in SAR images, the PSA-G module is proposed, combining AGTs with dynamic spatial attention mechanisms. This approach suppresses invalid noise interference while preserving detailed features, achieving synergistic optimization between feature decoupling and dynamic calibration, further enhancing the model's adaptability to complex backgrounds.

In the feature reconstruction and upsampling stage, the DySample module is introduced. Compared to traditional upsampling methods, DySample employs a "single-point dynamic dispersion into multiple points" mechanism, focusing more on the fine structures of small target regions and avoiding common issues such as feature loss and inefficient learning, thereby significantly improving feature restoration quality. The MSTFNet module adopts a cross-layer feature fusion strategy, integrating adjacent layer features while enhancing information flow between different semantic levels. This allows high-level abstract semantics to propagate to low-level detailed information, substantially improving the model's detection capability for small targets and spatial localization accuracy. Experimental results demonstrate that PPDM-YOLO achieves leading performance on two public datasets, HRSID and SSDD. On HRSID, it attains 93.7% mAP50 and 70.3% mAP50–95, while on SSDD, it reaches 99.4% and 78.7%, respectively. Compared to YOLOv11n, the model reduces parameters by 34.7% while surpassing it in accuracy, fully validating its comprehensive advantages in lightweight design, precision, and robustness.

Visualization results show that PPDM-YOLO delivers strong detection performance across various complex maritime backgrounds, particularly excelling in scenarios with densely distributed small targets. Notably, although the model has significantly reduced overall computational costs, there remains room for optimization in memory usage for extremely resource-constrained edge computing devices.

Future research will focus on further reducing model complexity and deployment costs. We plan to incorporate lightweight techniques such as model distillation and neural network pruning to compress and optimize the model structure, making it suitable for a broader range of real-time applications. In addition, we will explore multisource SAR image training, cross-platform deployment adaptation, and end-to-end joint training mechanisms for noise suppression and detection, aiming to provide more efficient and robust solutions for SAR image intelligent processing.

VI. Conclusion

Currently, a significant portion of research focuses on improving experimental accuracy through large-scale models, often overlooking the lightweight requirements associated with deployment. SAR ship imagery is characterized by complex backgrounds due to speckle noise and the presence of small, densely clustered targets. To address these challenges, we propose PPDM-YOLO, a lightweight SAR ship detection algorithm designed for complex environments. By integrating PCA, PSA-G, MSTFNet, and DySample modules, PPDM-YOLO not only demonstrates robust feature extraction capabilities but also effectively fuses multiscale features, excelling particularly in detecting densely distributed small objects. Compared to existing methods such as YOLOv11n, our approach achieves higher accuracy while significantly reducing the number of parameters. Extensive experiments on two benchmark datasets validate the effectiveness of the PCA, PSA-G, MSTFNet, and DySample components. PPDM-YOLO is well-suited for deployment on hardware across diverse and complex SAR ship detection scenarios.

References

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[2] A. Javali, J. Gupta, and A. Sahoo, "A review on synthetic aperture radar for earth remote sensing: Challenges and opportunities," in *Proc. 2nd Int. Conf. Electron. Sustain. Commun. Syst.*, 2021, pp. 596–601.

[3] D. R. Laura et al., "Synthetic Aperture Radar (SAR) image processing for operational space-based agriculture mapping," *Int. J. Remote Sens.* vol. 41, no. 18, pp. 7112–7144, 2020.

[4] Z. Zhao et al., "Ship surveillance by integration of space-borne SAR and AIS–review of current research," *J. Navigation*, vol. 67, no. 1, pp. 177–189, 2014.

[5] B. Hou, X. Chen, and L. Jiao, "Multilayer CFAR detection of ship targets in very high resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 811–815, Apr. 2015.

[6] T. Zeng et al., "CFAR-DP-FW: A CFAR-guided dual-polarization fusion framework for large-scene SAR ship detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 7242–7259, 2024.

[7] T. Li, Z. Liu, R. Xie, and L. Ran, "An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 184–194, Jan. 2018.

[8] X. Yang, X. Zhang, N. Wang, and X. Gao, "A robust one-stage detector for multiscale ship detection with complex background in massive SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5217712.

[9] Y. Song, J. Li, P. Gao, L. Li, T. Tian, and J. Tian, "Two-stage cross-modality transfer learning method for military-civilian SAR ship recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4506405.

[10] P. Jiang et al., "A review of YOLO algorithm developments," *Procedia Comput. Sci.*, vol. 199, pp. 1066–1073, 2022.

[11] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Comput. Vis.–ECCV 2016: 14th Eur. Conf., Amsterdam, The Netherlands, Oct. 11–14, 2016, Proc., Part I 14*, Springer Int. Publishing, 2016, pp. 21–37.

[12] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.

[13] R. Girshick, "Fast r-CNN[C]," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.

[14] S. Ren, K. He, R. Girshick, and J. Sun, "R-CNN Faster: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, 1137–1149, Jun. 2017.

[15] Y. Zhao, L. Zhao, B. Xiong, and G. Kuang, "Attention receptive pyramid network for ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2738–2756, 2020.

[16] Z. Wang, G. Hou, Z. Xin, G. Liao, P. Huang, and Y. Tai, "Detection of SAR image multiscale ship targets in complex inshore scenes based on improved YOLOv5," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 5804–5823, 2024.

[17] H. Zhu et al., "DB-YOLO: A duplicate bilateral YOLO network for multi-scale ship detection in SAR images," *Sensors*, vol. 21, no. 23, 2021, Art. no. 8146.

[18] J. Si, B. Song, J. Wu, W. Lin, W. Huang, and S. Chen, "Maritime ship detection method for satellite images based on multiscale feature fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 6642–6655, 2023.

[19] Z. Liangjun, N. Feng, X. Yubin, L. Gang, H. Zhongliang, and Z. Yuanyang, "MSFA-YOLO: A multi-scale SAR ship detection algorithm based on fused attention," *IEEE Access*, vol. 12, pp. 24554–24568, 2024.

[20] P. Huang, S. Tian, Y. Su, W. Tan, Y. Dong, and W. Xu, "IA-CIOU: An improved IOU bounding box loss function for SAR ship target detection methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 10569–10582, 2024.

[21] Y. Wang, G. Dong, S. Shao, and H. Liu, "A prediction and prior information guided SAR ship detection method," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7076–7088, 2023.

[22] Z. Hong et al., "Multi-scale ship detection from SAR and optical imagery via a more accurate YOLOv3," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6083–6101, 2021.

[23] X. Zhang, S. Feng, C. Zhao, Z. Sun, S. Zhang, and K. Ji, "MGSFA-Net: Multiscale global scattering feature association network for SAR ship target recognition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 4611–4625, 2024.

[24] Y. Gong, Z. Zhang, J. Wen, G. Lan, and S. Xiao, "Small ship detection of SAR images based on optimized feature pyramid and sample augmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7385–7392, 2023.

[25] X. Liu, J. Pan, R. Hu, W. Huang, J. Lin, and J. Hu, "DSMF-Net: A one-stage SAR ship detection network based on deformable strip convolution and multi-scale feature refinement and fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 10694–10710, 2025.

[26] Z. Sun et al., "Arbitrary-direction SAR ship detection method for multi-scale imbalance," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5208921.

[27] Y. Feng et al., "A lightweight position-enhanced anchor-free algorithm for SAR ship detection," *Remote Sens.*, vol. 14, no. 8, 2022, Art. no. 1908.

[28] Q. Mao, Y. Li, and Y. Zhu, "A hierarchical feature fusion and attention network for automatic ship detection from SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 13981–13994, 2024.

[29] B. Xiong et al., "A lightweight model for ship detection and recognition in complex-scene SAR images," *Remote Sens.*, vol. 14, no. 23, 2022, Art. no. 6053.

[30] Y. Guo et al., "LMSD-YOLO: A lightweight YOLO algorithm for multi-scale SAR ship detection," *Remote Sens.*, vol. 14, no. 19, 2022, Art. no. 4801.

[31] Y. Yin, X. Cheng, F. Shi, M. Zhao, G. Li, and S. Chen, "An enhanced lightweight convolutional neural network for ship detection in maritime surveillance system," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5811–5825, 2022.

[32] J. Cui, H. Jia, H. Wang, and F. Xu, "A fast threshold neural network for ship detection in large-scene SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6016–6032, 2022.

[33] P. Li and C. Che, "SeMo-YOLO: A multiscale object detection network in satellite remote sensing images," in *Proc. 2021 Int. Joint Conf. Neural Netw.*, Shenzhen, China, Jul. 2021, pp. 1–8.

[34] Y. Hao and Y. Zhang, "A lightweight convolutional neural network for ship target detection in SAR images," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 2, pp. 1882–1898, Apr. 2024.

[35] X. Xu, X. Zhang, and T. Zhang, "Lite-YOLOV5: A lightweight deep learning detector for on-board ship detection in large-scene sentinel-1 SAR images," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 1018.

[36] X. Tang, K. Cao, Y. Xia, E. Cui, W. Zhao, and Q. Chen, "BESW-YOLO: A lightweight SAR image detection model based on YOLOv8n for complex scenarios," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 16081–16094, 2025.

[37] S. Man and W Yu, "ELSD-Net: A novel efficient and lightweight ship detection network for SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 25, 2025, Art. no. 4003505.

[38] Y. Hao, J. Wu, Y. Yao, and Y. Guo, "A robust anchor-free detection method for SAR ship targets with lightweight CNN," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 5028319.

[39] G. Tang et al., "N-YOLO: A SAR ship detection using noise-classifying and complete-target extraction," *Remote Sens.*, vol. 13, no. 5, 2021, Art. no. 871.
[40] C. Zhao, X. Fu, J. Dong, S. Cao, and C. Zhang, "MLC-Net: A robust SAR ship detector with speckle noise and multiscale targets," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 19260–19273, 2024.
[41] C. Zhao, X. Fu, J. Dong, R. Qin, J. Chang, and P. Lang, "SAR ship detection based on end-to-end morphological feature pyramid network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4599–4611, 2022.
[42] Y. Dai, M. Zou, Y. Li, X. Li, K. Ni, and J. Yang, "Denodet: Attention as deformable multi-subspace feature denoising for target detection in SAR images," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 61, no. 2, pp. 4729–4743, Apr. 2025.
[43] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.
[44] T. Zhang et al., "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3690.
[45] R. Khanam and M. Hussain, "YOLOv11: An overview of the key architectural enhancements," 2024, *arXiv:2410.17725*.
[46] J. Chen et al., "Run, don't walk: Chasing higher FLOPS for faster neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 12021–12031.
[47] Q. Wang, B. Wu, and P. Zhu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11534–11542.
[48] W. Liu, H. Lu, H. Fu, and Z. Cao, "Learning to upsample by learning to sample," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 6027–6037.
[49] J. Wang, K. Chen, R. Xu, Z. Liu, C. C. Loy, and D. Lin, "Carafe: Content-aware reassembly of features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 3007–3016.
[50] M. M. Rahman, M. Munir, and R. Marculescu, "EMCAD: Efficient multi-scale convolutional attention decoding for medical image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 11769–11779.
[51] H. Lu et al., "SAPA: Similarity-aware point affiliation for feature upsampling," *Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 20889–20901, 2022.
[52] G. Jocher, A. Stoken, and J. Borovec, "Ultralytics/YOLOv5: v3. 0," *Zenodo*, 2020.
[53] M. Sohan et al., "A review on YOLOv8 and its advancements," in *Proc. Int. Conf. Data Intell. Cogn. Informat.*, Singapore. Springer, 2024, pp. 529–545.
[54] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
[55] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13713–13722.
[56] S. Woo et al., "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

Hongjie He received the Ph.D. degree in radio physics from Xidian University, Xi'an, China, in 2019.

He is currently an Associate Professor with the School of Electronic Information, Zhengzhou University of Light Industry, Zhengzhou, China. His research interests include computational electromagnetics and object detection and recognition.

Tianwen Hu received the B.Sc. degree in electronic information engineering in 2023 from Zhengzhou University of Light Industry, Zhengzhou, China, where he is currently working toward the M.S. degree focusing on SAR ship image target detection and deep learning.

His main research interests include SAR object detection and deep learning.

Sheng Xu received the Ph.D. degree from Xidian University, Xi'an, China, in 2014.

Since 2015, he has been with Zhengzhou University of Light Industry, Zhengzhou, China. His current research interests include ionospheric research and modeling and object detection and recognition

Hong Xu received the Ph.D. degree in control science and engineering from the Ordnance Engineering College, Shijiazhuang, China, in 2011.

Since 2017, he has been teaching with Zhengzhou University of Light Industry, Zhengzhou, China. His research interests mainly include wireless detection and positioning, intelligent sensing, and signal processing.

Lin Song received the B.Sc. degree in computer science and technology in 2022 from Zhengzhou University of Light Industry, Zhengzhou, China, where he is currently working toward the M.S. degree, focusing on remote sensing image target detection and deep learning.

His main research interests include object detection and deep learning.

Zhongpei Sun received the B.Sc. degree in electronic information science and technology in 2023 from Zhengzhou University of Light Industry, Zhengzhou, China, where he is currently working toward the M.S. degree, focusing on traffic vehicle target detection and deep learning.

His main research interests include deep learning-based object detection and multimodal fusion.