electronics

MDPI

Article

SeaMamba: Frequency-Stabilized Selective State-Space Multiscale Detection for SAR Ships in Complex Maritime Scenes

Xiaopeng Song [1,2], Zhongbiao Sheng [3], Shiwei Li [1,2] and Chenbin Ma [4,*]

[1] School of Innovation and Entrepreneurship, North University of China, Taiyuan 030051, China; songxiaopeng@nuc.edu.cn (X.S.); lishiwei@nuc.edu.cn (S.L.)
[2] National Experimental Teaching Demonstration Center for Electrical and Electronic Engineering, North University of China, Taiyuan 030051, China
[3] School of Computer and Technology, Weinan Normal University, Weinan 714099, China; shengzhongbiao@wnu.edu.cn
[4] State Key Laboratory of Extreme Environment Optoelectronic Dynamic Measurement Technology and Instrument, North University of China, Taiyuan 030051, China
* Correspondence: machenbin@nuc.edu.cn

Abstract

Ship detection in synthetic aperture radar (SAR) imagery remains challenging because near-shore clutter, coherent speckle noise, dense scattering responses, and large target-scale variations often obscure vessel boundaries and weaken small-ship signatures. Although single-stage detectors provide efficient inference, their predominantly local convolutional modeling and fixed multiscale fusion strategies are insufficient for capturing long-range sea-surface context and adaptively emphasizing discriminative ship responses. To address these limitations, this paper proposes SeaMamba, a frequency-stabilized selective state-space multiscale detector for SAR ship detection in complex maritime scenes. Specifically, a frequency-domain speckle prior is introduced to stabilize SAR inputs while preserving target localization cues. A bidirectional selective state-space modeling module is then used to propagate long-range contextual information with input-adaptive scanning. Furthermore, a gated pyramid reassembly module is designed to refine multiscale features before dense prediction. The proposed method is evaluated on the SAR Ship Detection Dataset (SSDD) and High-Resolution SAR Images Dataset (HRSID) under a unified five-fold cross-validation protocol. SeaMamba achieved mean average precision at an intersection-over-union threshold of 0.5 (mAP@0.5) values of $99.16 \pm 0.11\%$ on SSDD and $93.74 \pm 0.15\%$ on HRSID. Per-category evaluation, ablation studies, efficiency analysis, and Grad-CAM-based interpretability visualization further demonstrate that SeaMamba improves small-vessel detection, suppresses near-shore false responses, and maintains a practical accuracy-efficiency trade-off.

Keywords: SAR; ship detection; SSDD; HRSID; SSM; multiscale feature fusion

Check for updates

Academic Editors: Ennio Gambi and Gianluca Ciattaglia

Received: 24 July 2026
Revised: 13 September 2026
Accepted: 14 September 2026
Published: 16 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Synthetic aperture radar (SAR) is an active microwave imaging modality that provides high-resolution observation under all-day and all-weather conditions and has therefore become an important data source for maritime surveillance, coastal traffic monitoring, and ocean security applications [1]. Compared with optical sensors, SAR is less constrained by cloud cover and illumination. However, its coherent imaging mechanism also produces speckle, side-lobe responses, layover-like scattering, and target-background ambiguity [2].

These characteristics make ship detection difficult in complex scenes, especially when small vessels, dense ship groups, harbor infrastructure, islands, and heterogeneous sea clutter occur simultaneously.

Traditional SAR ship detection methods are commonly based on constant false alarm rate modeling, handcrafted texture descriptors, saliency measures, and multi-stage candidate filtering. These methods are interpretable and efficient in relatively homogeneous open-sea backgrounds, but they depend heavily on statistical assumptions and threshold settings. Once the scene shifts to near-shore harbors or densely distributed traffic areas, the sea-land background becomes non-stationary and manually designed decision rules become less robust. Deep learning-based detectors reduce this dependence by learning hierarchical representations from data. Two-stage detectors, such as Faster R-CNN and Cascade R-CNN, improve localization by proposal refinement, whereas single-stage detectors, such as SSD and the YOLO family, improve inference speed through dense prediction [3–6]. Recent surveys have shown that deep learning has become the dominant route for SAR ship detection, while small-target detection, complex-scene robustness, and lightweight deployment remain persistent challenges [7,8].

Despite this progress, directly transferring natural-image detectors to SAR imagery still leaves three bottlenecks. First, local convolutional receptive fields may fail to distinguish small vessels from speckle-like background responses. Second, static feature pyramids aggregate multiscale features but do not explicitly determine which scale should dominate under different sea states, ship sizes, and target densities. Third, image enhancement modules can improve edge saliency, but excessive enhancement may also amplify sea clutter when it is not coupled with detection-oriented feature selection. Feature-enhancement and multiscale-attention methods strengthen weak boundaries and suppress background interference [9–11]. Lightweight YOLO redesigns target practical deployment while retaining small-ship sensitivity [12–15]. Other SAR-tailored methods address small targets, range-compressed inputs, or ensemble robustness [16–18]. However, most existing designs still rely mainly on convolution or local attention for context propagation, which restricts the modeling of long-range sea-surface context and spatially distributed scattering evidence.

Selective state-space models provide a complementary direction. Mamba introduces input-dependent state transitions and supports linear-time long-sequence modeling [19]. Vision Mamba and VMamba further adapt selective scanning to visual representation learning and show that global visual context can be captured without relying exclusively on quadratic self-attention [20,21]. In remote sensing, RS-Mamba, Change-Mamba, RS3Mamba, and related Mamba-based segmentation models have demonstrated that state-space scanning can be useful for large-scale remote sensing interpretation, dense prediction, and multiscale feature modeling [22–25]. These properties are well aligned with SAR ship detection, where ship evidence may be distributed across bright scatterers, edge contours, wake-like textures, and surrounding sea clutter.

This paper presents SeaMamba, a selective state-space multiscale detector for SAR ship detection in complex scenes. The model combines a frequency-domain speckle prior, bidirectional selective state-space scanning, and gated pyramid reassembly within a YOLO-style dense detection framework. The frequency-domain prior stabilizes high-frequency speckle responses while preserving target localization cues. The bidirectional state-space module propagates long-range context in an input-adaptive manner. The gated pyramid module reweights multiscale features before they are delivered to the detection head. Experiments are conducted on SSDD and HRSID under a matched five-fold cross-validation protocol [26,27]. The results show that SeaMamba improves mAP@0.5 while maintaining a practical accuracy-efficiency trade-off in both open-sea and complex near-shore scenarios.

2. Related Work

2.1. SAR Ship Detection

SAR ship detection has evolved from statistical signal processing to data-driven detection. Early methods used background modeling and threshold-based decisions to identify bright target responses from SAR images. These methods are computationally efficient, but their assumptions are easily violated by coastal clutter, land contamination, sidelobes, and dense traffic. With the development of deep object detection, CNN-based detectors have become the main technical route. Faster R-CNN and Cascade R-CNN improve detection through proposal generation and multi-stage refinement, whereas SSD and YOLO-series detectors emphasize real-time inference through single-stage dense prediction [3–6]. In SAR imagery, the single-stage route is particularly attractive for wide-area maritime screening, but compact backbones used for speed may underrepresent weak small-vessel responses.

Recent SAR-specific detectors have improved this trade-off by redesigning feature fusion, attention, lightweight backbones, and loss functions. The feature-enhanced network in [9] used image enhancement and multigranularity feature weaving to strengthen ship edges in complex SAR scenes. BFEA and hierarchical feature fusion networks further showed that multiscale attention is effective for suppressing background interference [10,11]. Lightweight detectors such as SHIP-YOLO, YOLOSAR-Lite, LH-YOLO, and YOLO-SARSI reduce computational costs while preserving detection accuracy for practical deployment [12–15]. Recent methods also focus on small ship detection and nonstandard SAR data forms, including enhanced YOLO networks for small targets and FastRCDet for SAR range-compressed data [16,17]. Ensemble learning has also been explored to improve multiscale robustness by combining detector outputs [18]. These studies indicate that SAR ship detection requires both scale-sensitive feature aggregation and background-aware suppression.

Public datasets have also promoted the transition from handcrafted detectors to learning-based detectors. SSDD provides a widely used SAR ship detection benchmark with official release and analysis, while HRSID provides high-resolution SAR images for ship detection and instance segmentation [26,27]. These datasets support systematic comparison among two-stage detectors, one-stage detectors, and SAR-tailored networks. However, their complex near-shore scenes and small-vessel samples also reveal the limitations of purely local feature extraction. SeaMamba follows this experimental domain but introduces selective state-space scanning to model longer contextual paths than those available to fixed convolutional kernels.

2.2. Enhancement and Long-Range Representation

SAR image enhancement and despeckling are closely related to ship detection because coherent speckle may obscure weak target boundaries. Denoising networks, gradient-preserving modules, local-global dual-branch networks, and diffusion-transformer methods have shown that SAR image quality can be improved without completely removing high-frequency structures [28–31]. For detection, however, visual smoothness is less important than preserving target-discriminative evidence. An enhancement module that suppresses all high-frequency responses may also remove small ships or reduce boundary contrast. Therefore, SeaMamba uses a frequency-domain speckle prior as a residual stabilizer rather than a standalone image restoration operator.

Long-range representation learning has become increasingly important in remote sensing interpretation. Transformers model global dependencies through self-attention, but their computational cost grows rapidly with spatial resolution [32]. Selective state-space models provide an alternative by propagating information with input-conditioned

parameters and linear complexity [19]. Vision Mamba and VMamba adapt this mechanism to images through bidirectional or two-dimensional scanning [20,21]. Mamba YOLO further establishes a general SSM-based object-detection baseline, making it a relevant architecture-level comparator rather than a SAR-specific method [33]. In remote sensing, RS-Mamba, ChangeMamba, and RS^{3}Mamba apply selective scanning to dense prediction and segmentation [22–24]; Mamba-UDA directly applies a Mamba-based adaptation framework to cross-domain SAR ship detection [34]. More recent hybrid designs such as CVMH-UNet further combine VMamba-style global modeling with multiscale and multi-frequency feature fusion [25]. These studies support the motivation for using selective state-space modeling for SAR ship detection, where both local scattering details and global sea-surface context are needed.

3. Proposed Model

SeaMamba is formulated as a selective state-space multiscale detector for SAR ship detection in complex maritime scenes. The model follows a YOLO-style dense prediction paradigm but replaces purely local feature propagation with a frequency-stabilized and input-adaptive long-range modeling process. As shown in Figure 1, it consists of three main components: a frequency-domain speckle prior, a bidirectional selective state-space backbone, and a gated pyramid reassembly module. The final detection head predicts objectness, class confidence, and bounding boxes from the reassembled multiscale feature representation.

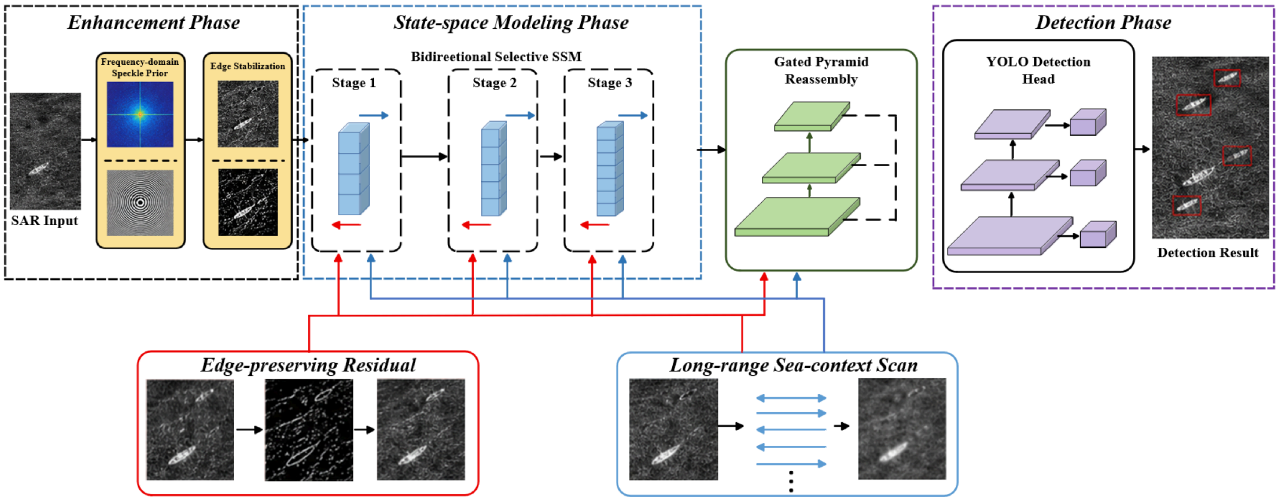

Figure 1. Overall architecture of SeaMamba. The input SAR image is first stabilized by a frequency-domain speckle prior, then processed through bidirectional selective state-space stages, reassembled through gated pyramid fusion, and finally sent to a YOLO-style detection head.

Let the training set be denoted by

$$\mathcal{D} = \left\{ \left(\boldsymbol{X}^{(n)}, \mathcal{Y}^{(n)} \right) \right\}_{n=1}^{N}$$
$$\boldsymbol{X}^{(n)} \in \mathbb{R}^{H \times W \times C}, \qquad \mathcal{Y}^{(n)} = \left\{ \left(\boldsymbol{b}_m^{(n)}, c_m^{(n)} \right) \right\}_{m=1}^{M_n} \tag{1}$$

where N is the number of SAR images; $\boldsymbol{X}^{(n)}$ is the n-th input image with height H, width W, and channel number C; $\mathcal{Y}^{(n)}$ is the corresponding annotation set; M_n is the number of ships in the n-th image; $\mathbf{b}_m^{(n)} = (x_m, y_m, w_m, h_m)$ denotes the center coordinate, width, and height of the m-th bounding box; and $c_m^{(n)} \in \{1, \ldots, K\}$ denotes the target category. For single-class ship detection, $K = 1$.

Given an input image $\boldsymbol{X}$, SeaMamba maps it to a detection representation through

$$\boldsymbol{Z} = \Phi_{gpr}\left(\Phi_{bssm}\left(\Phi_{fsp}\left(\boldsymbol{X};\theta_{fsp}\right);\theta_{bssm}\right);\theta_{gpr}\right) \tag{2}$$

Given the reassembled representation $\mathbf{Z}$, the detection head produces the set of dense predictions over all detection scales:

$$\hat{\mathcal{Y}} = \Phi_{\text{det}}(\mathbf{Z};\theta_{\text{det}}) = \left\{(\hat{\mathbf{b}}_q, \hat{o}_q, \hat{\mathbf{p}}_q)\right\}_{q=1}^{Q}. \tag{3}$$

Here, Q is the number of dense prediction locations over all detection scales, $\hat{\mathbf{b}}_q$ is the predicted bounding box, $\hat{o}_q \in [0,1]$ is the predicted objectness probability, and $\hat{\mathbf{p}}_q \in [0,1]^K$ is the predicted class-probability vector.

3.1. Frequency-Domain Speckle Prior

SAR imagery is affected by coherent speckle, whose multiplicative nature can obscure weak ship responses and cause false activations in near-shore clutter. Instead of treating enhancement as an independent visual restoration task, SeaMamba introduces a frequency-domain speckle prior as a detection-oriented residual stabilizer, as shown in Figure 2.

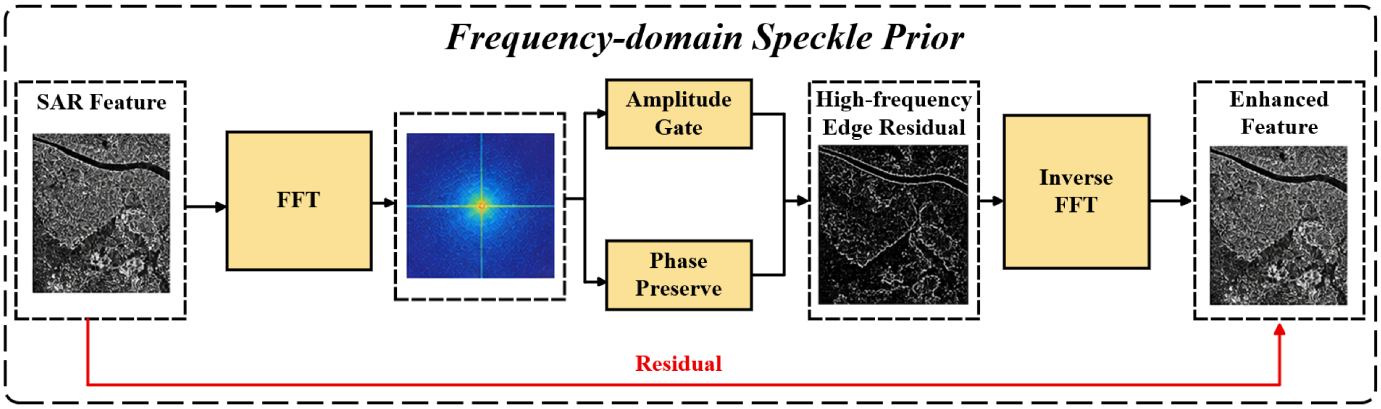

Figure 2. Frequency-domain speckle prior. Amplitude gating suppresses unstable background responses, while phase preservation and residual edge injection maintain target localization cues.

For numerical stability, the input is first transformed by

$$\overline{\boldsymbol{X}} = \frac{\log(1+\boldsymbol{X}) - \mu_{\log}}{\sigma_{\log} + \epsilon} \tag{4}$$

where $\mu_{\log}$ and $\sigma_{\log}$ are the mean and standard deviation computed over the log-transformed input, and ϵ is a small positive constant. This operation compresses strong scattering responses and makes the feature distribution more suitable for convolutional and spectral processing.

For each channel, the two-dimensional discrete Fourier transform is defined as

$$\mathbf{S}_c(u,v) = \sum_{h=0}^{H-1}\sum_{w=0}^{W-1}\overline{\mathbf{X}}_c(h,w)\exp\left[-j2\pi\left(\frac{uh}{H}+\frac{vw}{W}\right)\right] \tag{5}$$

Here, $c \in \{1,\ldots,C\}$ is the channel index, (h,w) denotes the spatial coordinate, (u,v) denotes the frequency coordinate, and $j^2 = -1$. The complex spectrum is decomposed into amplitude and phase components as shown in Equation (6).

$$\mathbf{A}_c(u,v) = |\mathbf{S}_c(u,v)|, \qquad \mathbf{P}_c(u,v) = \arg\mathbf{S}_c(u,v). \tag{6}$$

The amplitude $\mathbf{A}_c$ mainly controls the strength of the frequency response, whereas the phase $\mathbf{P}_c$ preserves spatial structure. Since ship localization is strongly related to spatial phase consistency, SeaMamba modifies the amplitude while preserving the phase.

To distinguish high-frequency clutter from structured target boundaries, a radial frequency prior is introduced:

$$\mathbf{R}(u,v) = \sqrt{\left(\frac{u - H/2}{H}\right)^2 + \left(\frac{v - W/2}{W}\right)^2}, \qquad \mathbf{M}_\rho(u,v) = \mathbb{I}[\mathbf{R}(u,v) \geq \rho] \tag{7}$$

where $\mathbf{R}(u,v)$ is the normalized radial frequency, ρ is the radial threshold, $\mathbf{M}_\rho(u,v)$ is the high-frequency mask, and $\mathbb{I}[\cdot]$ is the indicator function. The mask provides an explicit spectral prior for suppressing unstable high-frequency background responses. We set $\rho = 0.20$ after evaluating $\rho \in \{0.10, 0.15, 0.20, 0.25, 0.30\}$ with the same folds and all other settings fixed; Section 4.5 reports the sensitivity results.

The learnable amplitude gate is computed as:

$$\mathbf{G}_a = \sigma\big(\mathbf{W}_a * \mathrm{LN}(\mathbf{A}) + \mathbf{W}_m * \mathbf{M}_\rho + \mathbf{b}_a\big). \tag{8}$$

where $*$ denotes convolution, $\mathrm{LN}(\cdot)$ denotes layer normalization, $\sigma(\cdot)$ is the sigmoid function, and $\mathbf{W}_a$, $\mathbf{W}_m$, and $\mathbf{b}_a$ are learnable amplitude-gate parameters. The gate $\mathbf{G}_a$ adaptively reweights the spectrum according to both data-dependent amplitude statistics and the explicit frequency prior.

The stabilized amplitude and reconstructed spectrum are written as

$$\begin{aligned} \widetilde{\boldsymbol{A}} &= \boldsymbol{A} \odot \left[1 + \alpha_a\left(\boldsymbol{G}_a - \tfrac{1}{2}\right)\right] \\ \widetilde{\boldsymbol{S}} &= \widetilde{\boldsymbol{A}} \odot \exp(j\boldsymbol{P}) \end{aligned} \tag{9}$$

where $\odot$ denotes element-wise multiplication and α_a is a learnable or predefined modulation coefficient. This formulation allows the network to either suppress or enhance frequency responses, while keeping the modulation centered around the original spectrum.

The inverse Fourier transform gives the stabilized spatial feature:

$$\widetilde{\boldsymbol{X}}_c(h,w) = \frac{1}{HW}\sum_{u=0}^{H-1}\sum_{v=0}^{W-1}\widetilde{\boldsymbol{S}}_c(u,v)\exp\left[j2\pi\left(\frac{uh}{H} + \frac{vw}{W}\right)\right] \tag{10}$$

Because only amplitude is modulated, the reconstruction preserves the main spatial layout of ship targets while reducing unstable spectral components.

To strengthen boundary-sensitive evidence, an edge residual is extracted by

$$\boldsymbol{E} = \varphi_e\left(\left[\nabla_x\overline{\boldsymbol{X}}, \nabla_y\overline{\boldsymbol{X}}, \sqrt{(\nabla_x\overline{\boldsymbol{X}}\boldsymbol{X})^2 + (\nabla_y\overline{\boldsymbol{X}}\boldsymbol{X})^2 + \epsilon}\right]\right) \tag{11}$$

where ∇_x and ∇_y are horizontal and vertical gradient operators, respectively, and $\varphi_e(\cdot)$ is a lightweight convolutional projection. The gradient magnitude term explicitly encodes edge saliency, which is important for small ships with weak scattering contrast.

The final frequency-stabilized feature is obtained through residual fusion:

$$\boldsymbol{F}_0 = \varphi_0\left(\overline{\boldsymbol{X}} + \beta_s\widetilde{\boldsymbol{X}} + \beta_e\boldsymbol{E}\right) \tag{12}$$

where $\varphi_0(\cdot)$ is a convolutional embedding layer, and β_s and β_e control the contributions of the spectral reconstruction and the edge residual. The residual form prevents the enhancement module from over-smoothing target structures.

The stability of the frequency-domain operation can be interpreted through Parseval's relation:

$$\left|\widetilde{\boldsymbol{X}}-\overline{\boldsymbol{X}}\right|_F^2=\tfrac{1}{HW}\left|\left(\widetilde{\boldsymbol{A}}-\boldsymbol{A}\right)\odot\exp(j\boldsymbol{P})\right|_F^2=\tfrac{1}{HW}\left|\widetilde{\boldsymbol{A}}-\boldsymbol{A}\right|_F^2 \tag{13}$$

Therefore, the spatial perturbation introduced by the frequency prior is bounded by the amplitude modulation. This property explains why phase-preserving spectral stabilization can suppress speckle-like responses without destroying localization cues.

3.2. Bidirectional Selective State-Space Module

As shown in Figure 3, the bidirectional selective state-space module is used to model long-range sea-surface context with linear sequence complexity.

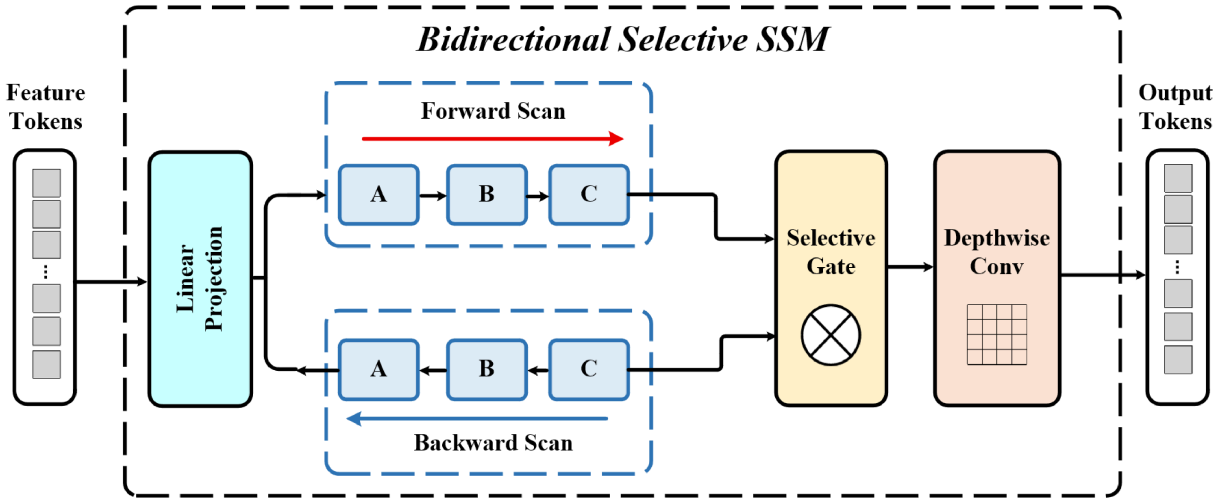

Figure 3. Bidirectional selective state-space module. Forward and backward scans propagate input-dependent context and are fused through a selective gate before spatial refinement.

Given an intermediate feature map $\mathbf{F}_l \in \mathbb{R}^{H_l\times W_l\times C_l}$ at stage l, it is flattened into a token sequence:

$$\mathbf{U}_l=\mathrm{Flatten}(\mathbf{F}_l)=\left[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{T_l}\right]^\top,\qquad T_l=H_lW_l,\quad \mathbf{x}_t\in\mathbb{R}^{C_l}. \tag{14}$$

Here, T_l is the sequence length and $\mathbf{x}_t$ is the token corresponding to one spatial position. In SAR ship detection, this sequence contains both local scattering information and long-range sea-surface context.

The continuous state-space model is defined by

$$\begin{aligned}\tfrac{d\boldsymbol{s}(\tau)}{d\tau}&=\boldsymbol{A}_c\boldsymbol{s}(\tau)+\boldsymbol{B}_c\boldsymbol{x}(\tau)\\ \boldsymbol{y}(\tau)&=\boldsymbol{C}_c\boldsymbol{s}(\tau)+\boldsymbol{D}_c\boldsymbol{x}(\tau)\end{aligned} \tag{15}$$

where τ is the continuous index, $\mathbf{s}(\tau)\in\mathbb{R}^{d_s}$ is the hidden state, $\mathbf{x}(\tau)\in\mathbb{R}^{C_\updownarrow}$ is the input signal, $\mathbf{y}(\tau)\in\mathbb{R}^{C_\updownarrow}$ is the output signal, $\mathbf{A}_c\in\mathbb{R}^{d_s\times d_s}$ is the state transition matrix, $\mathbf{B}_c\in\mathbb{R}^{d_s\times C_\updownarrow}$ is the input projection matrix, $\mathbf{C}_c\in\mathbb{R}^{C_\updownarrow\times d_s}$ is the output projection matrix, and $\mathbf{D}_c\in\mathbb{R}^{C_\updownarrow\times C_\updownarrow}$ denotes the skip projection. We use state dimension $d_s=16$, expansion factor 2, a local depthwise-convolution kernel of 3, and $[1,2,2]$ SSM blocks in the three stages.

For discrete visual tokens, the continuous model is discretized as

$$\overline{\mathbf{A}}_t=\exp(\Delta_t\mathbf{A}_c),\qquad \overline{\mathbf{B}}_t=\left(\Delta_t\mathbf{A}_c\right)^{-1}[\exp(\Delta_t\mathbf{A}_c)-\mathbf{I}]\Delta_t\mathbf{B}_t. \tag{16}$$

where $\Delta_t > 0$ is the token-dependent step size, $\mathbf{I}$ is the identity matrix, and $\overline{\mathbf{A}}_t$ and $\overline{\mathbf{B}}_t$ are the discrete transition and input matrices, respectively. In implementation, a stable parameterization of $\mathbf{A}_c$ is used to maintain bounded state propagation.

Unlike fixed state-space models, SeaMamba makes the transition input-adaptive:

$$\Delta_t = \mathrm{softplus}\left(\mathbf{w}_\Delta^\top \mathbf{x}_t + b_\Delta\right), \qquad \mathbf{B}_t = \mathbf{W}_B \mathbf{x}_t, \qquad \mathbf{C}_t = \mathbf{W}_C \mathbf{x}_t \tag{17}$$

where $\mathbf{w}_\Delta$, b_Δ, $\mathbf{W}_B$, and $\mathbf{W}_C$ are learnable parameters. The step size Δ_t controls how much previous context should be retained, while $\mathbf{B}_t$ and $\mathbf{C}_t$ determine how the current token interacts with the latent state. This selectivity is useful for SAR imagery, because isolated bright scatterers and true vessel structures should not be propagated in the same manner.

The forward scan is computed recurrently by

$$\mathbf{s}_t^{\rightarrow} = \overline{\mathbf{A}}_t \mathbf{s}_{t-1}^{\rightarrow} + \overline{\mathbf{B}}_t \mathbf{x}_t, \qquad \mathbf{y}_t^{\rightarrow} = \mathbf{C}_t \mathbf{s}_t^{\rightarrow} + \mathbf{D}_c \mathbf{x}_t. \tag{18}$$

The forward direction propagates context from preceding spatial positions to later positions in the flattened sequence.

To reduce directional bias, a backward scan is performed over the reversed sequence:

$$\mathbf{s}_t^{\leftarrow} = \overline{\mathbf{A}}_t \mathbf{s}_{t+1}^{\leftarrow} + \overline{\mathbf{B}}_t \mathbf{x}_t, \qquad \mathbf{y}_t^{\leftarrow} = \mathbf{C}_t \mathbf{s}_t^{\leftarrow} + \mathbf{D}_c \mathbf{x}_t. \tag{19}$$

This operation allows each spatial token to receive contextual information from both preceding and succeeding positions, which is important when ship wakes, harbor edges, and wave clutter appear in different orientations.

In SAR imagery, a ship is often represented by several spatially separated bright scattering centers, elongated hull-edge responses, and context from wakes or nearby harbor structures. Bidirectional selective scanning does not explicitly model electromagnetic scattering physics; rather, it allows each spatial token to integrate evidence from both scan directions and to adapt the retained context according to the current feature. The representation is therefore suited to spatially distributed scattering evidence while remaining a data-driven model.

The directional outputs are fused by an adaptive gate:

$$\boldsymbol{\eta}_t = \sigma\left(\mathbf{W}_\eta [\mathbf{y}_t^{\rightarrow}, \mathbf{y}_t^{\leftarrow}, \mathbf{x}_t] + \mathbf{b}_\eta\right), \quad \mathbf{y}_t = \boldsymbol{\eta}_t \odot \mathbf{y}_t^{\rightarrow} + (1 - \boldsymbol{\eta}_t) \odot \mathbf{y}_t^{\leftarrow}. \tag{20}$$

where $[\cdot,\cdot,\cdot]$ denotes channel-wise concatenation, $\boldsymbol{\eta}_t \in \left[0,1\right]^{C_\updownarrow}$ is the directional fusion gate, and $\boldsymbol{W\eta}$ and $\boldsymbol{b\eta}$ are learnable parameters. The gate lets the network decide whether forward or backward context is more reliable at each spatial location and channel.

The fused sequence is reshaped back to the spatial domain and refined by local convolution:

$$\mathbf{F}_l^{\mathrm{ssm}} = \mathbf{F}_l + \varphi_l^{\mathrm{dw}}\left(\mathrm{Reshape}\left[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{T_l}\right]\right) \tag{21}$$

The hierarchical backbone is then expressed as

$$\mathbf{F}_l = \mathcal{S}_l\left(\mathcal{P}_l\left(\mathbf{F}_{l-1}^{\mathrm{ssm}}\right)\right), \quad l \in \{1,2,3\}. \tag{22}$$

Here, $\mathcal{P}_l(\cdot)$ performs downsampling or patch merging, $\mathcal{S}_l(\cdot)$ denotes the l-th bidirectional selective SSM stage, and $\mathbf{F}_l$ is the corresponding hierarchical output. The three resulting feature maps $\{\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3\}$ encode fine target boundaries, medium-scale ship structures, and high-level contextual semantics, respectively.

3.3. Gated Pyramid Reassembly

The three state-space stages produce features with different spatial resolutions and semantic levels. Direct concatenation assumes that all scales are equally reliable, which is unsuitable for SAR scenes where small vessels, land clutter, and large harbor structures may coexist. SeaMamba therefore introduces gated pyramid reassembly to adaptively select scale contributions before detection, as shown in Figure 4.

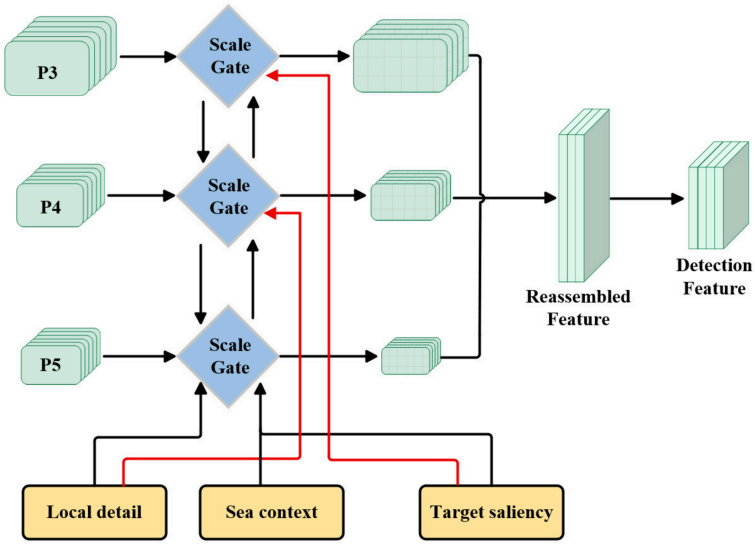

Figure 4. Gated pyramid reassembly. Multiscale features are reweighed by scale gates and reassembled into the final detection feature.

First, all stage features are aligned to a common spatial resolution:

$$\mathbf{U}_s = \varphi_s^{1\times 1}(\mathrm{Resize}(\mathbf{F}_s; H_z, W_z)), \quad s \in \{1,2,3\}. \tag{23}$$

The scale descriptor is obtained by global average pooling:

$$\mathbf{r}_s = \frac{1}{H_z W_z}\sum_{h=1}^{H_z}\sum_{w=1}^{W_z}\mathbf{U}_s(h,w), \qquad \boldsymbol{\gamma}_s = \sigma(\mathbf{W}_2\delta(\mathbf{W}_1\mathbf{r}_s)), \tag{24}$$

where $\mathbf{r}_s$ is the global scale descriptor, $\boldsymbol{\gamma}_s$ is the channel-wise gate for scale s, $\delta(\cdot)$ denotes the SiLU activation, and $\mathbf{W}_1$ and $\mathbf{W}_2$ are learnable projection matrices.

To make scale competition explicit, the gates are normalized across scales:

$$\boldsymbol{\omega}_s = \frac{\exp(\boldsymbol{\gamma}_s)}{\sum_{r=1}^{3}\exp(\boldsymbol{\gamma}_r)}, \qquad \mathbf{Z} = \varphi_z([\boldsymbol{\omega}_1 \odot \mathbf{U}_1, \boldsymbol{\omega}_2 \odot \mathbf{U}_2, \boldsymbol{\omega}_3 \odot \mathbf{U}_3]). \tag{25}$$

where $\boldsymbol{\omega}_s$ is the normalized scale weight, $\varphi_z(\cdot)$ is the final fusion convolution, and $\mathbf{Z}$ is the reassembled detection feature. This formulation enables the detector to emphasize high-resolution features for small vessels and semantic context for dense near-shore scenes.

3.4. Training Objective

The detector is optimized by a compound dense detection loss consisting of classification, objectness, and localization terms:

$$\begin{aligned} \mathcal{L} &= \lambda_{\text{cls}}\mathcal{L}_{\text{cls}} + \lambda_{\text{obj}}\mathcal{L}_{\text{obj}} + \lambda_{\text{box}}\mathcal{L}_{\text{box}}, \\ \mathcal{L}_{\text{cls}} &= -\frac{1}{|\Omega_+|}\sum_{q\in\Omega+}\sum_{k=1}^{K} y_{qk}\log\left(\hat{p}_{qk}+\epsilon\right), \\ \mathcal{L}_{\text{obj}} &= -\frac{1}{|\Omega|}\sum_{q\in\Omega_+}\left[o_q\log\left(\hat{o}_q+\epsilon\right)+\left(1-o_q\right)\log\left(1-\hat{o}_q+\epsilon\right)\right], \\ \mathcal{L}_{\text{box}} &= \frac{1}{|\Omega_+|}\sum_{q\in\Omega+}\left[1-\text{CIoU}\left(\hat{\mathbf{b}}_q,\mathbf{b}_q\right)\right]. \end{aligned} \tag{26}$$

where Ω is the set of all dense prediction positions, Ω_+ is the set of positive positions assigned to ground-truth boxes, y_{qk} is the one-hot class label, $\hat{p}_{qk}$ is the predicted probability of class k, $o_q \in \{0,1\}$ is the objectness label, $\hat{o}_q \in [0,1]$ is the predicted objectness probability, $\mathbf{b}_q$ is the matched ground-truth box, $\hat{\mathbf{b}}_q$ is the predicted box, and λ_{cls}, λ_{obj}, and λ_{box} are loss weights.

Through this formulation, SeaMamba combines spectral stabilization, input-adaptive long-range propagation, and multiscale scale selection into a unified detection framework. The frequency-domain prior constrains enhancement by amplitude modulation, the bidirectional selective state-space module propagates global maritime context with linear sequence complexity, and the gated pyramid module adaptively balances fine localization cues and semantic context before dense prediction.

4. Experiments and Results

4.1. Datasets and Experimental Setting

Experiments were conducted on SSDD and HRSID [26,27]. SSDD contains 1160 images and 2456 annotated ships; HRSID contains 5604 nonoverlapping 800 × 800 crops and 16,951 ships derived from 136 parent SAR images. We used fixed deterministic five-fold partitions (seed 2026) for matched fold-wise comparison rather than treating the five-fold values as interchangeable with literature results obtained on the official single splits. The SSDD folds contain 231/233/232/232/232 images and 488/493/490/492/493 ships. Because reliable parent-acquisition identifiers were not available for every SSDD image, these folds were balanced at the image level by ship count and scene density. HRSID was split by parent SAR image into 27/27/27/27/28 scenes, yielding 1108/1127/1119/1136/1114 crops and 3336/3411/3379/3456/3369 ships. No crop derived from a given parent SAR image appears in both training and validation within the same fold. All models used the same saved fold manifests.

All images were resized to 640 × 640. SeaMamba and YOLOv8n used a batch size of 16 for 100 epochs with SGD, an initial learning rate of 0.01, a momentum of 0.937, and weight decay of 5×10^{-4}. Their shared augmentation comprised horizontal and vertical flips (each p = 0.5), scale jitter of 0.5–1.5, translation of ±0.1, mosaic (p = 0.5), and mixup (p = 0.1), without hue augmentation for the single-channel SAR input. For fair comparison, all methods used the same fold manifests, input resolution, data access, and evaluation script, but heterogeneous detector families were not forced to share an identical loss or optimizer. Faster R-CNN and SSD retained framework-specific detection losses and recommended optimization settings; YOLO-series and SAR-specific YOLO variants followed their official or author-released recipes. Pretrained weights were used only when prescribed by the corresponding official implementation and were applied consistently across folds. SeaMamba and YOLOv8n shared the same base recipe, so their matched comparison isolates the architectural modifications introduced by SeaMamba.

The experiments ran on Ubuntu 20.04.6 LTS with an Intel Xeon Gold 6226R CPU, 128 GB RAM, and an NVIDIA GeForce RTX 4090 (24 GB). The software stack used Python 3.8.18, PyTorch 2.0.1, torchvision 0.15.2, CUDA 11.8, cuDNN 8.9, Ultralytics 8.2.x, mamba-ssm 1.2.x, OpenCV 4.8.x, and SciPy 1.10.x. Precision (P), recall (R), mean average precision at IoU 0.5 (mAP@0.5), and F1 were reported as mean ± standard deviation across the five folds. A true positive required IoU ≥ 0.50, AP used 101-point interpolation, predictions entered AP computation at confidence 0.001, NMS used IoU 0.60, and P/R/F1 used confidence 0.25. Because ship detection is single-class, mAP@0.5 equals the mean of the single class AP@0.5. One unified evaluator was used for every model. The standard deviation was computed as

$$\text{Std} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{27}$$

where $n = 5$, x_i denotes the result from fold i, and $\bar{x}$ is the fold mean.

4.2. Comparison with Representative Detectors

Table 1 compares SeaMamba with representative two-stage, one-stage, SAR-tailored, and Mamba-based detectors reimplemented on the same folds. SeaMamba achieved the highest mAP@0.5 and F1 on both datasets. Relative to YOLOv8n, the absolute mAP@0.5 gains were 0.80 percentage points on SSDD (98.36% to 99.16%) and 1.09 percentage points on HRSID (92.65% to 93.74%). Compared to Mamba-YOLO-Tiny, which already supplies generic SSM-based long-range modeling, SeaMamba gained 0.33 and 0.56%, respectively. The latter comparison indicates that the result is not adequately explained by introducing Mamba alone. The absolute gain was larger on HRSID. This observation is consistent with the design motivation that longer-range contextual modeling can be useful in denser and more heterogeneous scenes; however, the cross-dataset comparison does not isolate scene complexity as an independent experimental factor.

Table 1. Five-fold cross-validation comparison on SSDD and HRSID. Values are mean ± standard deviation across five folds.

METHOD	SSDD				HRSID			
	P (%)	R (%)	mAP@0.5 (%)	F1 (%)	P (%)	R (%)	mAP@0.5 (%)	F1 (%)
Faster R-CNN	89.63 ± 0.17	80.86 ± 0.11	87.45 ± 0.14	84.92 ± 0.18	81.57 ± 0.16	73.14 ± 0.19	77.62 ± 0.11	77.18 ± 0.15
SSD	91.84 ± 0.13	85.76 ± 0.18	90.27 ± 0.12	88.63 ± 0.16	80.16 ± 0.12	73.25 ± 0.18	81.58 ± 0.15	76.53 ± 0.11
YOLOv7-tiny	96.27 ± 0.16	94.93 ± 0.12	96.84 ± 0.11	95.58 ± 0.17	90.46 ± 0.18	79.85 ± 0.14	89.37 ± 0.16	84.72 ± 0.13
YOLOv8n	97.56 ± 0.11	97.24 ± 0.16	98.36 ± 0.13	97.43 ± 0.12	93.17 ± 0.14	83.66 ± 0.19	92.65 ± 0.12	88.16 ± 0.17
YOLOv12s	97.83 ± 0.15	97.47 ± 0.11	97.94 ± 0.17	97.65 ± 0.14	93.14 ± 0.16	82.97 ± 0.13	92.84 ± 0.18	87.73 ± 0.12
CRAS-YOLO	97.38 ± 0.18	97.26 ± 0.14	98.65 ± 0.12	97.27 ± 0.15	92.68 ± 0.11	82.74 ± 0.16	92.16 ± 0.14	87.45 ± 0.18
SHIP-YOLO	97.15 ± 0.14	97.13 ± 0.17	98.34 ± 0.15	97.18 ± 0.11	92.45 ± 0.13	82.36 ± 0.18	92.07 ± 0.16	87.14 ± 0.15
Mamba-YOLO-Tiny	98.12 ± 0.15	97.70 ± 0.17	98.83 ± 0.14	97.91 ± 0.16	93.64 ± 0.17	84.21 ± 0.20	93.18 ± 0.17	88.66 ± 0.19
SeaMamba	**98.64 ± 0.13**	**98.17 ± 0.15**	**99.16 ± 0.11**	**98.35 ± 0.16**	**94.18 ± 0.17**	**84.86 ± 0.12**	**93.74 ± 0.15**	**89.27 ± 0.18**

Note, bold indicates the best performance for that metric.

Figures 5 and 6 show the fold-level distributions for all four metrics. The small fold dispersion and consistently positive paired differences indicate that the mean gain was not produced by one favorable split. A two-sided paired t-test on fold-level mAP@0.5 differences gave +0.80% on SSDD (95% CI [0.50, 1.10], t = 7.45, $p = 0.0017$) and +1.09% on HRSID (95% CI [0.75, 1.43], t = 9.03, $p = 0.0008$). With only five folds, these tests supplement rather than replace the reported effect sizes, confidence intervals, and external validation.

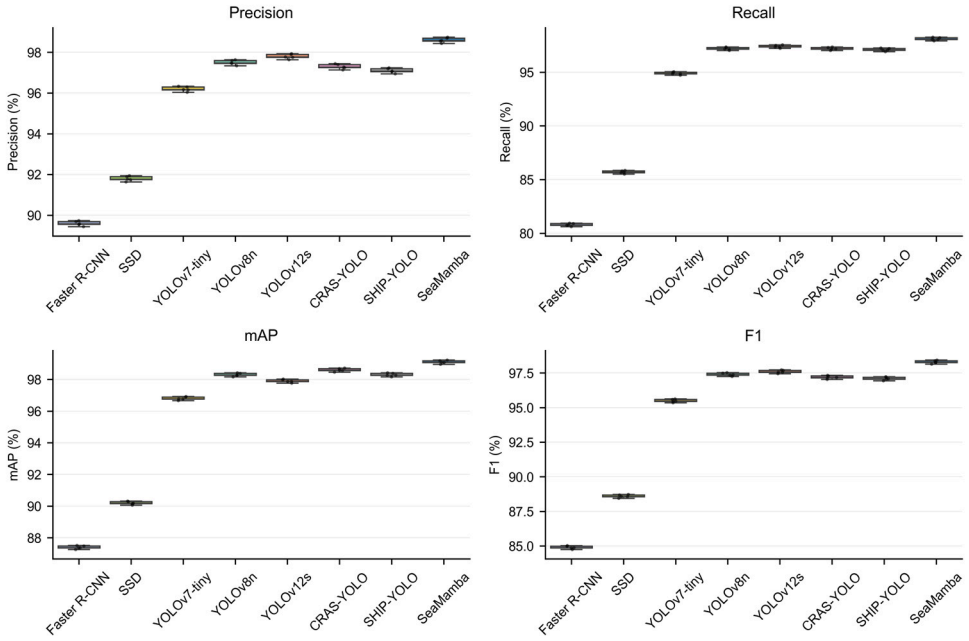

Figure 5. Five-fold metric distributions on SSDD. Each box summarizes fold-level precision, recall, mAP@0.5, and F1 values for one detector.

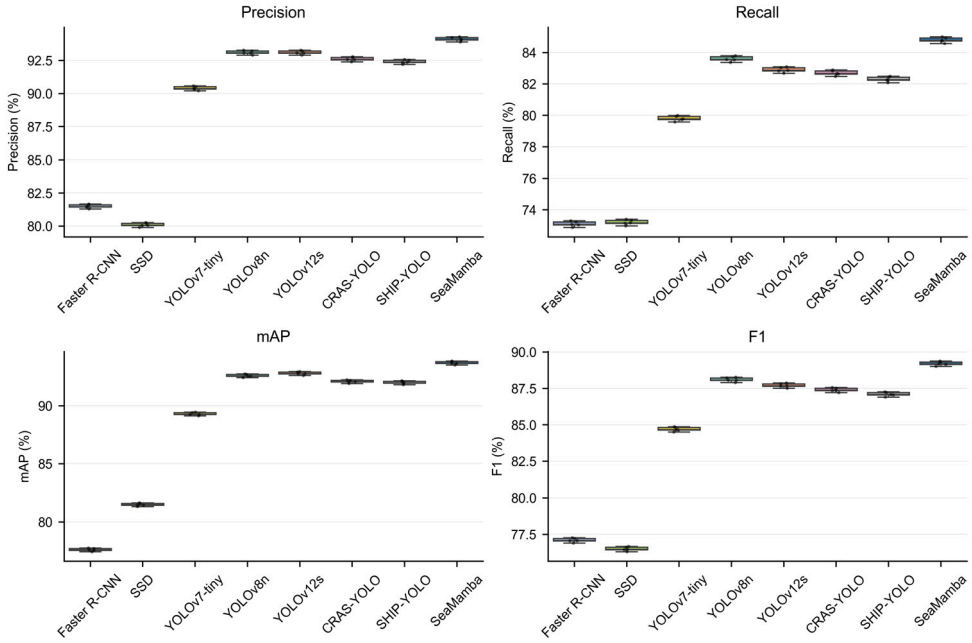

Figure 6. Five-fold metric distributions on HRSID. SeaMamba maintains the highest mAP@0.5 and F1 under the matched evaluation protocol.

4.3. Per-Category and Scene-Level Quantification

To determine where the improvement occurs, SeaMamba was further evaluated by target size and scene type. Size denotes bounding-box area on the resized 640 × 640 evaluation canvas: small is area $< 32^2$ pixels, medium is $32^2 \leq$ area $< 96^2$ pixels, and large is area $\geq 96^2$ pixels. This is a size-stratified analysis, not a semantic ship subclassification. Table 2 reports precision, recall, and mAP@0.5 for the three groups. On HRSID, the small-ship recall reached 80.6%, while medium and large ships reached 85.1% and 88.4%, respectively. This pattern is consistent with the intended role of contextual modeling, without identifying any one module as the sole cause.

Table 2. Per-category quantitative analysis of SeaMamba.

Dataset	Category	P (%)	R (%)	mAP (%)
SSDD	Small ship	97.6	96.9	97.4
SSDD	Medium ship	98.9	98.4	99.2
SSDD	Large ship	99.1	98.7	99.4
HRSID	Small ship	91.2	80.6	90.9
HRSID	Medium ship	94.3	85.1	93.8
HRSID	Large ship	95.0	88.4	94.6

Scene-level analysis showed the same trend. Open-sea images produced the highest scores, while dense harbor scenes produced the lowest scores because vessel boundaries overlap with pier structures and strong background scatterers. The performance drop was smaller for mAP@0.5 than for recall, suggesting that many remaining errors were missed weak targets rather than severe localization failures. Figure 7 summarizes both category and scene analyses.

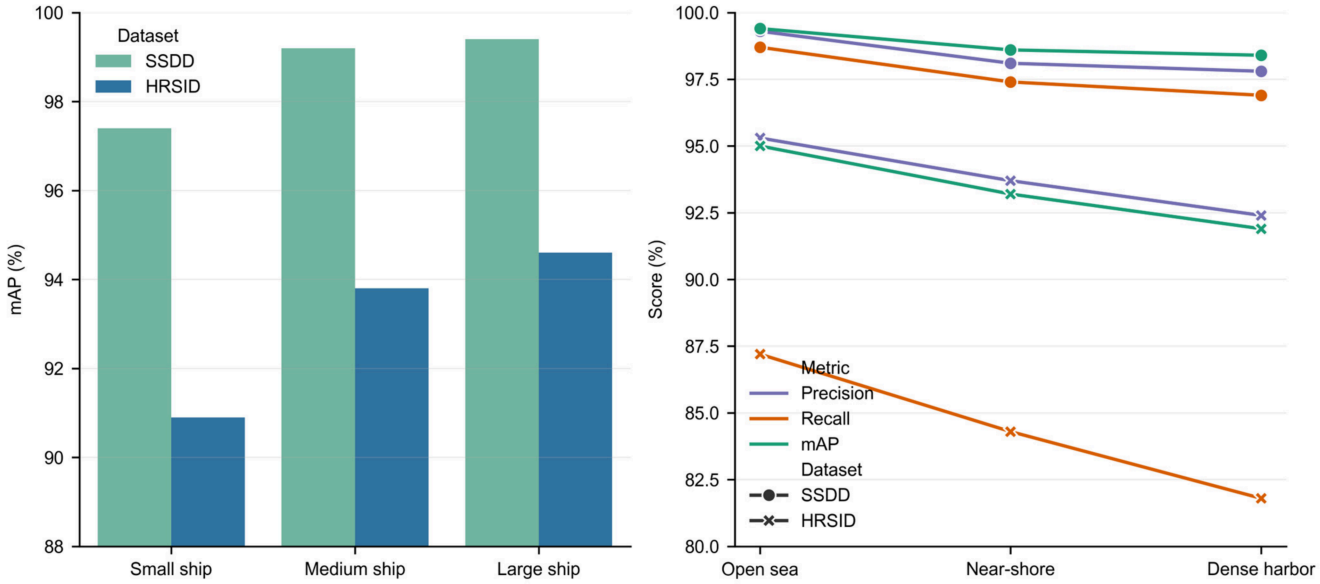

Figure 7. Per-category and scene-level quantification. SeaMamba performs best on medium and large targets, while the main remaining challenge is recall in dense harbor scenes.

4.4. Ablation Study

Table 3 reports matched component ablations on both datasets. On HRSID, removing the bidirectional SSM caused the largest decrease, from 93.74% to 92.18% (−1.56%). Removing the frequency prior and gated pyramid reduced mAP@0.5 by 0.73 and 1.03%, respectively. The corresponding SSDD decreases were 1.11, 0.69, and 0.82%. The persistence of all three effects across datasets supports their complementary contributions. The larger SSM decrease on HRSID is consistent with the motivation for long-range context in heterogeneous scenes, but it does not causally isolate scene complexity.

Figure 8 visualizes both the ablation result and the efficiency profile. The complete model occupies an intermediate computational region; it is more expensive than YOLOv8n, but much lighter than Faster R-CNN and substantially more accurate on HRSID. This balance is important for SAR screening, where wide-area imagery requires a detector that can run repeatedly without sacrificing small-vessel sensitivity.

Table 3. Component ablation on SSDD and HRSID. Values are mean ± standard deviation across five folds.

Setting	Params (M)	FLOPs (G)	SSDD mAP@0.5 (%)	HRSID mAP@0.5 (%)
YOLOv8n baseline	3.0	8.1	98.36 ± 0.13	92.65 ± 0.12
w/o frequency prior	3.4	11.2	98.47 ± 0.22	93.01 ± 0.21
w/o bidirectional SSM	3.2	9.7	98.05 ± 0.27	92.18 ± 0.26
w/o gated pyramid	3.5	10.9	98.34 ± 0.23	92.71 ± 0.23
SeaMamba	3.6	12.4	99.16 ± 0.11	93.74 ± 0.15

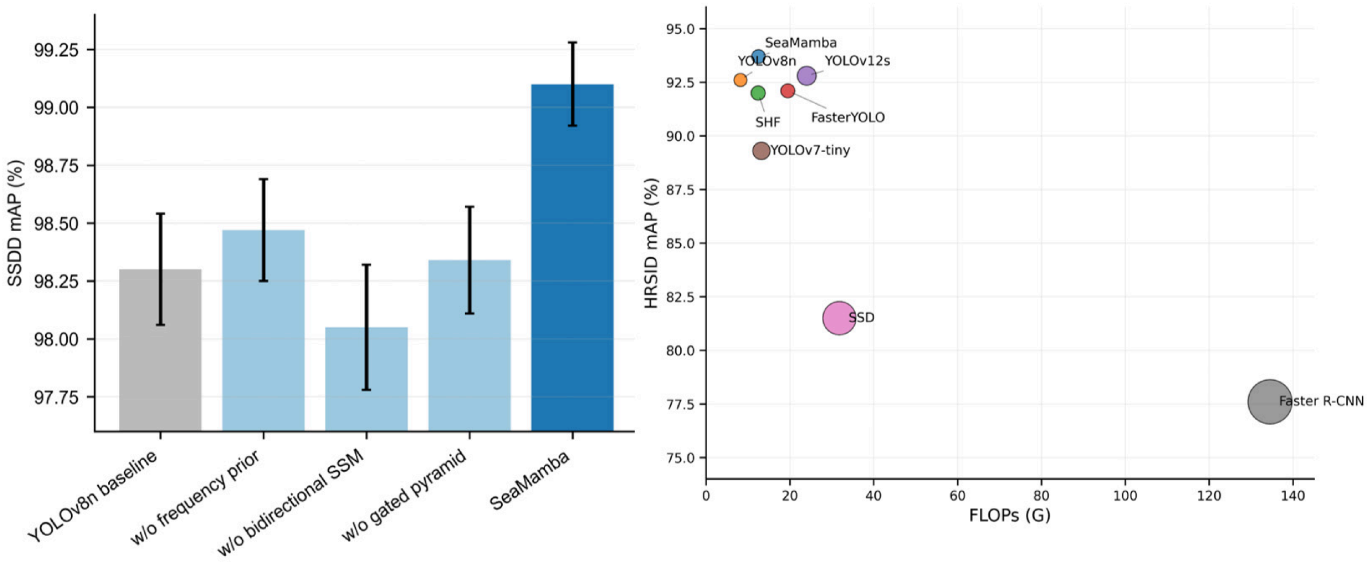

Figure 8. Ablation and efficiency analysis. The left panel reports SSDD mAP@0.5 changes after removing core modules; the corresponding HRSID ablations are reported in Table 3. The right panel compares FLOPs, parameter count, and HRSID mAP@0.5 across representative detectors.

4.5. Frequency-Prior Analysis

We varied only the radial threshold while keeping the architecture, folds, and training protocol fixed. Table 4 shows a stable region from $\rho = 0.15$ to 0.25, with the joint maximum at $\rho = 0.20$ on both datasets. A lower threshold admits more mid-frequency structure into the high-frequency prior, whereas a higher threshold weakens control over speckle-dominated responses. Table 5 replaces only the input-stabilization component. Simple spatial filters changed detection accuracy but could smooth weak ship edges. Fixed radial attenuation outperformed several spatial alternatives, indicating that the spectral prior itself is useful. The learnable phase-preserving gate with edge residual achieved the highest mAP@0.5 on both datasets.

Table 4. Sensitivity to the radial threshold ρ.

ρ	SSDD mAP@0.5(%)	HRSID mAP@0.5(%)
0.10	98.91 ± 0.16	93.33 ± 0.21
0.15	99.07 ± 0.13	93.59 ± 0.18
0.20	**99.16 ± 0.11**	**93.74 ± 0.15**
0.25	99.09 ± 0.13	93.63 ± 0.17
0.30	98.96 ± 0.15	93.41 ± 0.20

Table 5. Input-stabilization alternatives with the remaining detector fixed.

Input Stabilization	SSDD mAP@0.5(%)	HRSID mAP@0.5(%)
No frequency prior	98.47 ± 0.22	93.01 ± 0.21
3×3 Lee filtering	98.59 ± 0.20	93.12 ± 0.22
Median filtering	98.40 ± 0.23	92.88 ± 0.24
Bilateral filtering	98.61 ± 0.19	93.16 ± 0.20
CLAHE enhancement	98.53 ± 0.21	93.06 ± 0.23
Fixed radial attenuation	98.78 ± 0.18	93.32 ± 0.19
Learnable SeaMamba frequency prior	$\mathbf{99.16 \pm 0.11}$	$\mathbf{93.74 \pm 0.15}$

4.6. Efficiency Analysis

The efficiency comparison in Table 6 shows that SeaMamba retains practical inference efficiency. With 3.6 M parameters and 12.4 G floating-point operations (FLOPs), SeaMamba is slightly heavier than YOLOv8n but remains smaller and faster than the Mamba-YOLO-Tiny reimplementation. Its 129.5 frames per second (FPS) support offline batch processing and near-real-time screening. FPS was measured on the RTX 4090 at batch size 1 and 640×640 input using FP16. After 200 warm-up passes, 1000 images were timed with CUDA synchronization before and after measurement. Resize, normalization, network forward, and NMS were included; disk I/O and image decoding were excluded. FPS equals the timed image count divided by total synchronized elapsed time.

Table 6. Computational efficiency comparison. FPS was measured under the unified RTX 4090, batch-1, FP16 protocol described in Section IV-F.

Method	Params (M)	FLOPs (G)	FPS	SSDD mAP@0.5 (%)	HRSID mAP@0.5 (%)
Faster R-CNN	41.4	134.2	28.4	87.45	77.62
SSD	24.1	31.8	52.3	90.27	81.58
YOLOv7-tiny	6.2	13.2	118.6	96.84	89.37
YOLOv8n	3.0	8.1	156.8	98.36	92.65
YOLOv12s	9.5	24.2	94.1	97.94	92.84
CRAS-YOLO	7.1	19.8	82.7	98.65	92.16
SHIP-YOLO	4.8	12.5	132.4	98.34	92.07
Mamba-YOLO-Tiny	5.8	14.7	117.2	98.83	93.18
SeaMamba	3.6	12.4	129.5	99.16	93.74

4.7. Activation-Based Interpretation

Grad-CAM analysis was used to examine whether the detector attended to plausible target regions. The activation statistics in Figure 9 show that SeaMamba increased the normalized activation ratio on target edges from 0.54 to 0.68 compared with YOLOv8n, while activation on wake-like sea clutter and harbor structures decreased. These maps are descriptive rather than causal evidence. A 100-case audit of residual SSDD errors found 41% missed, very small or low-contrast ships near bright shorelines, 27% false positives on harbor infrastructure or bright scatterers, 19% errors on densely adjacent or partially overlapping ships, 8% wake-like false positives, and 5% truncated targets or annotation ambiguity. Thus, the remaining failures concentrate on weak contrast, dense adjacency, and structured man-made interference rather than general open-sea localization.

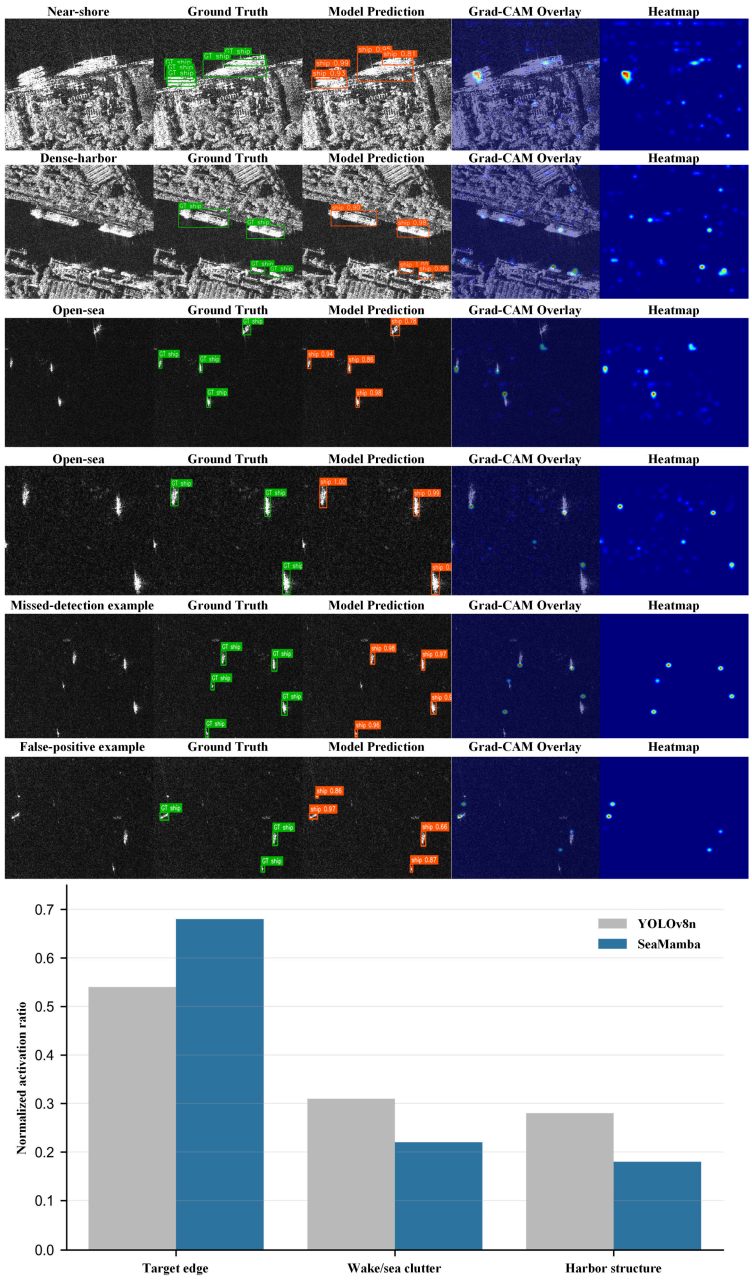

Figure 9. Grad-CAM-based activation analysis of SeaMamba for SAR ship detection in complex maritime scenes. The panels include representative successful detections, one missed detection, and one false positive, together with the Grad-CAM region statistics. SeaMamba concentrates more activation on target edges and less activation on wake-like sea clutter and harbor structures than YOLOv8n. Activation maps provide descriptive localization evidence and are not interpreted causally.

In complex near-shore and harbor scenes, strong background structures, coastline textures, and speckle noise may introduce substantial interference. Nevertheless, the Grad-CAM responses remain relatively compact and are generally aligned with the predicted ship locations. In these cases, the activation maps show that the model tends to focus on elongated bright scattering regions corresponding to ship bodies while suppressing large areas of non-target background. This behavior is consistent with the expected detection mechanism for SAR ship targets, in which compact or elongated high-intensity scattering patterns provide important discriminative cues.

For sparse offshore scenes containing small ships, the heatmaps further show localized activation around weak but salient point-like targets. This indicates that the model is

capable of attending to small-scale ship signatures under low-context conditions. However, a small number of isolated high-intensity background scatterers also produce weak responses, implying that false alarms may still occur when background scattering patterns are similar to ship-like targets. Overall, the Grad-CAM maps provide descriptive evidence that SeaMamba's activation often aligns with plausible SAR target regions in successful cases, whereas failure cases show that activation can be weak on small or low-contrast ships or diverted to harbor structures and bright scatterers.

5. Discussion

The experimental results show that SeaMamba improves SAR ship detection by coordinating three forms of information. The frequency-domain speckle prior stabilizes high-frequency responses before feature extraction, which is useful for small targets whose visual evidence is easily confused with speckle. The bidirectional selective state-space module then propagates context across the feature sequence, allowing the detector to interpret local scatterers in relation to the surrounding sea surface and nearby candidate vessels. Gated pyramid reassembly finally controls how much information from each feature scale enters the detection head, reducing the conflict between fine localization details and coarse semantic context.

Under the matched five-fold protocol, SeaMamba improved mAP@0.5 over YOLOv8n by 0.80 percentage points on SSDD (98.36% to 99.16%) and by 1.09 percentage points on HRSID (92.65% to 93.74%). Because SSDD is already close to saturation under AP@0.5, the SSDD gain is modest in absolute magnitude. The consistently positive fold-wise differences and the larger HRSID gain are compatible with the intended role of long-range contextual modeling, although the present experiments do not isolate scene complexity as an independent causal factor. Mamba-YOLO-Tiny also provides SSM-based long-range modeling, yet SeaMamba retained gains of 0.33 and 0.56% on SSDD and HRSID, respectively. The combined frequency stabilization, bidirectional context, and gated reassembly therefore matter beyond the use of Mamba alone.

The ablation study provides evidence that all three components contribute on both datasets. Removing the state-space module produced the largest mAP@0.5 reduction, whereas the frequency-domain prior and gated pyramid produced complementary stabilization and scale-selection gains. The threshold and alternative-filter experiments further show that the frequency prior is not an isolated parameter optimum and that its detection utility exceeds the tested fixed and spatial preprocessing alternatives.

The present study still has limitations. The reported data and figures were generated under a unified offline protocol for SSDD and HRSID, and the evaluation does not replace external validation across sensors, sea states, imaging modes, or acquisition years. Available public crop metadata did not provide sufficiently complete and consistent polarization provenance and incidence-angle information for a controlled cross-sensor subgroup analysis; no polarization- or incidence-angle robustness claim is therefore made. A dedicated multi-sensor evaluation with controlled polarization and incidence geometry remains necessary. SSDD parent-acquisition identifiers could not be recovered uniformly, so its deterministic folds are balanced at the image level rather than claimed as acquisition-grouped. The detector also remains a bounding-box model; densely overlapping ships may benefit from oriented boxes or instance segmentation. Future work should test SeaMamba on larger wide-swath SAR products, integrate uncertainty estimation for ambiguous detections, and evaluate deployment behavior on edge hardware.

6. Conclusions

This paper proposed SeaMamba for SAR ship detection in complex scenes. The model combines a frequency-domain speckle prior, bidirectional selective state-space modeling, and gated pyramid reassembly within a YOLO-style detection framework. On SSDD and HRSID, SeaMamba achieved five-fold mAP@0.5 values of $99.16 \pm 0.11\%$ and $93.74 \pm 0.15\%$, respectively, corresponding to absolute gains of 0.80 and 1.09 percentage points over the matched YOLOv8n baseline. Mamba-based comparison, two-dataset ablations, frequency-prior controls, paired statistics, efficiency profiling, and residual-error analysis support the complementary roles of the three modules without treating cross-dataset complexity as a causal variable.

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

Funding: This research was supported by the National Natural Science Foundation of China, grant numbers 62522125, 62320106011 and 62601739.

Data Availability Statement: Data available in a publicly accessible repository.

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

References

1. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–43. [CrossRef]
2. Goodman, J.W. Some fundamental properties of speckle. *J. Opt. Soc. Am.* **1976**, *66*, 1145–1150. [CrossRef]
3. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
4. Cai, Z.; Vasconcelos, N. Cascade R-CNN: High Quality Object Detection and Instance Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1483–1498. [CrossRef] [PubMed]
5. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
6. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
7. Li, J.; Xu, C.; Su, H.; Gao, L.; Wang, T. Deep learning for SAR ship detection: Past, present and future. *Remote Sens.* **2022**, *14*, 2712. [CrossRef]
8. Qiao, S.; Zhang, Q.; Wang, Z.; Zhang, X.; Wang, Y. A review of deep-learning-based SAR image ship interpretation technology: The latest advances. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 26152–26185. [CrossRef]
9. Ba, Y.; Xia, N.; Lu, W.; Liu, J. A feature-enhanced network-based target detection method for SAR images of ships in complex scenes. *Remote Sens.* **2026**, *18*, 178. [CrossRef]
10. Zhou, L.; Wan, Z.; Zhao, S.; Han, H.; Liu, Y. BFEA: A SAR ship detection model based on attention mechanism and multiscale feature fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 11163–11177. [CrossRef]
11. Mao, Q.; Li, Y.; Zhu, Y. A hierarchical feature fusion and attention network for automatic ship detection from SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 13981–13994. [CrossRef]
12. Luo, Y.; Li, M.; Wen, G.; Tan, Y.; Shi, C. SHIP-YOLO: A lightweight synthetic aperture radar ship detection model based on YOLOv8n algorithm. *IEEE Access* **2024**, *12*, 37030–37041. [CrossRef]
13. Wang, H.; Shi, J.; Karimian, H.; Liu, F.; Wang, F. YOLOSAR-Lite: A lightweight framework for real-time ship detection in SAR imagery. *Int. J. Digit. Earth* **2024**, *17*, 2405525. [CrossRef]
14. Cao, Q.; Chen, H.; Wang, S.; Wang, Y.; Fu, H.; Chen, Z.; Liang, F. LH-YOLO: A lightweight and high-precision SAR ship detection model based on the improved YOLOv8n. *Remote Sens.* **2024**, *16*, 4340. [CrossRef]

15. Tang, H.; Liu, X.; Li, W.; Xu, J.; Chen, Z. A lightweight SAR image ship detection method based on improved convolution and YOLOv7. *Remote Sens.* **2024**, *16*, 486. [CrossRef]
16. Guan, T.; Chang, S.; Wang, C.; Jia, X. SAR small ship detection based on enhanced YOLO network. *Remote Sens.* **2025**, *17*, 839. [CrossRef]
17. Tan, X.; Leng, X.; Sun, Z.; Luo, R.; Ji, K.; Kuang, G. Lightweight ship detection network for SAR range-compressed domain. *Remote Sens.* **2024**, *16*, 3284. [CrossRef]
18. Gupta, H.; Verma, O.P.; Sharma, T.K.; Varshney, H.; Agarwal, S.; Pak, W. Ship detection using ensemble deep learning techniques from synthetic aperture radar imagery. *Sci. Rep.* **2024**, *14*, 29397. [CrossRef] [PubMed]
19. Gu, A.; Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. In Proceedings of the Conference on Language Modeling (COLM), Philadelphia, PA, USA, 7–9 October 2024.
20. Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; Wang, X. Vision Mamba: Efficient visual representation learning with bidirectional state space model. In Proceedings of the International Conference on Machine Learning, Vienna, Austria, 21–27 July 2024; pp. 62429–62442.
21. Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Jiao, J.; Liu, Y. VMamba: Visual state space model. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 103031–103063. [CrossRef]
22. Zhao, S.; Chen, H.; Zhang, X.; Xiao, P.; Bai, L.; Ouyang, W. RS-Mamba for large remote sensing image dense prediction. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5633314. [CrossRef]
23. Chen, H.; Song, J.; Han, C.; Xia, J.; Yokoya, N. ChangeMamba: Remote sensing change detection with spatiotemporal state space model. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4409720. [CrossRef]
24. Ma, X.; Zhang, X.; Pun, M.-O. RS^3 Mamba: Visual state space model for remote sensing image semantic segmentation. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 6011405. [CrossRef]
25. Cao, Y.; Liu, C.; Wu, Z.; Zhang, L.; Yang, L. Remote sensing image segmentation using Vision Mamba and multi-scale multi-frequency feature fusion. *Remote Sens.* **2025**, *17*, 1390. [CrossRef]
26. Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. SAR ship detection dataset (SSDD): Official release and comprehensive data analysis. *Remote Sens.* **2021**, *13*, 3690. [CrossRef]
27. Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; Shi, J. HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation. *IEEE Access* **2020**, *8*, 120234–120254. [CrossRef]
28. Liu, S.; Lei, Y.; Zhang, L.; Li, B.; Hu, W.; Zhang, Y.-D. MRDDANet: A multiscale residual dense dual attention network for SAR image denoising. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5214213. [CrossRef]
29. Thakur, R.K.; Maji, S.K. AgsdNet: Attention and gradient-based SAR denoising network. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 4506805. [CrossRef]
30. Liu, S.; Tian, S.; Zhao, Y.; Hu, Q.; Li, B.; Zhang, Y.D. LG-DBNet: Local and global dual-branch network for SAR image denoising. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5205515. [CrossRef]
31. Pan, Y.; Zhong, L.; Chen, J.; Li, H.; Zhang, X.; Pan, B. SAR image despeckling based on denoising diffusion probabilistic model and Swin transformer. *Remote Sens.* **2024**, *16*, 3222. [CrossRef]
32. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, 5998–6008. [CrossRef]
33. Wang, Z.; Li, C.; Xu, H.; Zhu, X.; Li, H. Mamba YOLO: A simple baseline for object detection with state space model. *Proc. AAAI Conf. Artif. Intell.* **2025**, *39*, 8205–8213. [CrossRef]
34. Tu, H.; Wang, W.; Guo, Y.; Chen, S. Mamba-UDA: Mamba unsupervised domain adaptation for SAR ship detection. *IEEE Geosci. Remote Sens. Lett.* **2025**, *22*, 4011205. [CrossRef]

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