remote sensing

MDPI

Article

MSPaDet: A Multi-Scale Phase-Aware Denoising Method for Target Detection in SAR Images

Naxiong Chen, Xuyu Xiang * and Yuanjing Luo

College of Computer and Mathematics, Central South University of Forestry and Technology, Changsha 410004, China; chennaxiong@csuft.edu.cn (N.C.); luoyuanjing@csuft.edu.cn (Y.L.)
* Correspondence: xyuxiang@csuft.edu.cn

Highlights

What are the main findings?

- Explicit phase-coherent multi-scale frequency decomposition provides a principled mechanism for separating weak SAR targets from clutter, as it suppresses low-coherence speckle responses while preserving structurally meaningful directional cues.
- Joint modeling of phase and energy is shown to yield a more discriminative and structurally stable representation than single-cue enhancement, enabling more faithful characterization of weak, irregular, and anisotropically scattered targets under complex background interference.

What are the implication of the main findings?

- The results suggest that SAR target detection can benefit from moving beyond purely spatial enhancement or coarse magnitude-based frequency selection toward explicit phase-aware, direction-sensitive, and multi-scale representation learning.
- By improving the structural fidelity and stability of target representations under speckle, clutter, and anisotropic scattering, the proposed strategy offers a practically meaningful route toward more reliable SAR perception in Earth observation, maritime surveillance, and safety-critical monitoring applications.

Abstract

Synthetic Aperture Radar (SAR) target detection remains challenging due to coherent speckle corruption, weak-scattering targets with degraded structural cues, and cross-scale inconsistencies under anisotropic scattering. To tackle these challenges, this paper presents MSPaDet, a novel multi-scale phase-aware denoising detection framework that advances SAR target detection by deeply integrating phase coherence with multi-scale representation learning. The proposed method introduces explicit dual-tree complex wavelet transform decomposition to generate direction-selective complex sub-bands, enabling fine-grained sub-band modulation. Within the framework, an SCFRDeno module suppresses speckle-dominant responses while preserving high-frequency structures via phase-coherence-guided reweighting, and a PaSCA block further refines features through input-adaptive spatial focusing and region reweighting. Extensive experiments on public SAR detection benchmarks—including MSAR, SAR-Aircraft-1.0, and SARDet-100K—demonstrate that our approach consistently outperforms state-of-the-art methods in detection accuracy, robustness, and cross-scenario generalization, with moderate computational cost, showing promising potential for practical deployment in Earth observation and safety monitoring systems.

Check for updates

Academic Editor: Dusan Gleich

Received: 17 March 2026
Revised: 7 May 2026
Accepted: 8 May 2026
Published: 11 May 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.

Keywords: synthetic aperture radar; target detection; speckle noise suppression; dual-tree complex wavelet transform; phase coherence

1. Introduction

Synthetic Aperture Radar (SAR) is an active microwave imaging sensor that enables all-weather observation by penetrating clouds, rain, and fog [1]. Reliable detection and localization of ships, vehicles, and critical infrastructure in complex land–sea backgrounds are crucial for Earth observation applications and safety-critical decision-making [2]. However, as shown in Figure 1, real-world SAR targets are often small, weak, and irregular, and are heavily affected by speckle and clutter. Coherent speckle degrades texture cues and target–background contrast [3,4], while small and weak targets provide limited cues that are easily submerged by noise [5,6]. Moreover, anisotropic scattering induces cross-scale response instability, impairing multi-scale fusion and generalization.

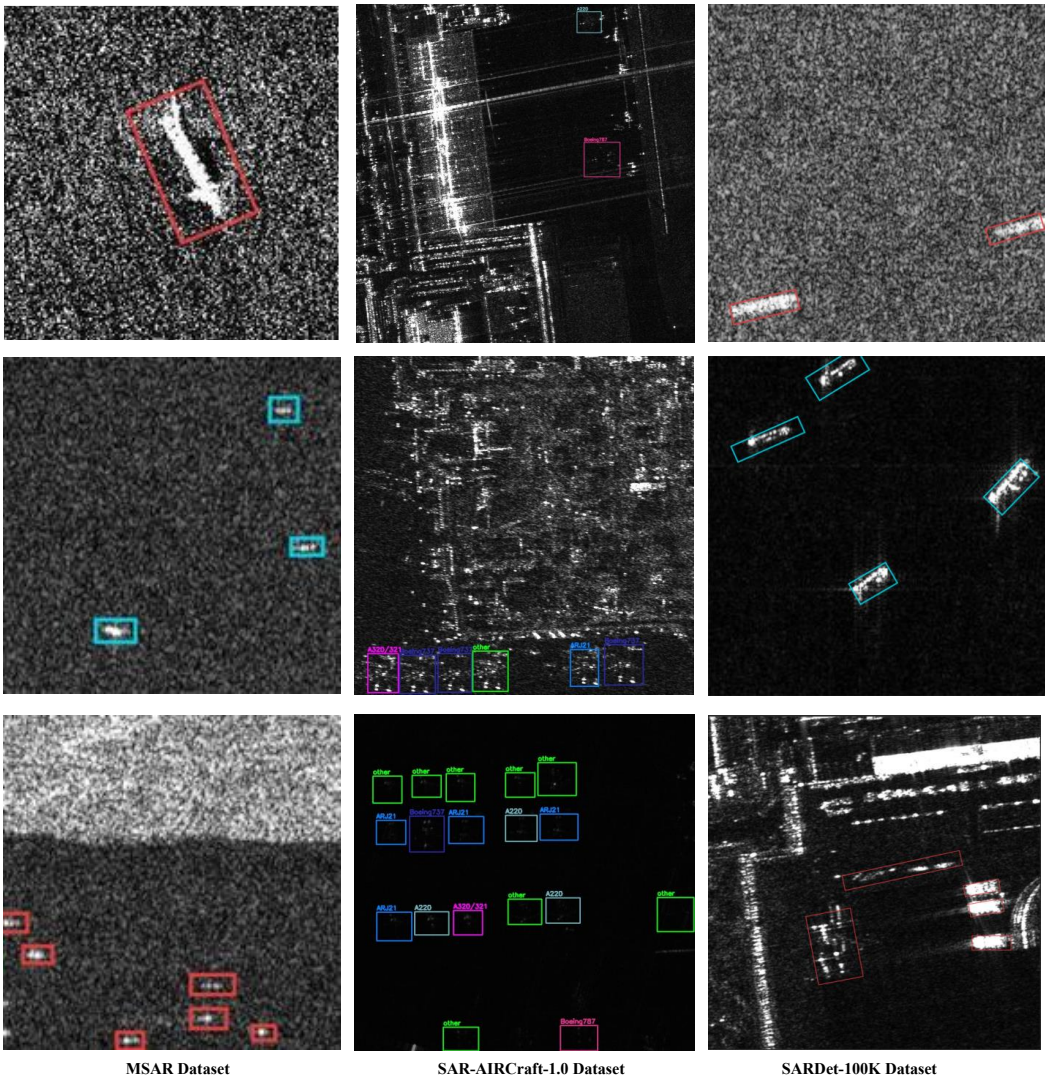


Figure 1. Representative images for remote SAR target detection.

Image resolution directly affects target visibility and target–background separability in SAR target detection [7]. In practical spaceborne SAR systems, high-resolution imaging is commonly achieved through spotlight or sliding spotlight modes. Spotlight SAR provides finer azimuth resolution for detailed observation, whereas sliding spotlight SAR offers a

practical trade-off between high resolution and wider azimuth coverage. Although these acquisition modes improve structural detail representation, they also make speckle, clutter, and weak-target detection more challenging, which further motivates the need for robust feature modeling in SAR target detection [8,9].

Before the rise of deep learning, SAR detection largely relied on clutter modeling and threshold-based schemes such as CFAR, often assisted by saliency analysis, clustering, or polarization information to improve detectability [1,2,10,11]. Recent deep learning methods have gradually addressed the problems of heavy speckle, small targets, irregular targets, and scale instability by embedding denoising and detail-preserving modules into detectors or integrating conventional frequency attention (as illustrated in Figure 2), strengthening multi-scale fusion and training strategies, and introducing consistency constraints and adaptive mechanisms to mitigate performance fluctuations caused by resolution variations [12–18]. SAR ship detection is also an important branch of SAR target detection, especially in maritime surveillance and large-scene ocean observation. Recent studies, such as LS-SSDD-v1.0 for small-ship detection and scale-aware dimension-wise attention networks for SAR small-ship instance segmentation, further highlight that small object scale, weak contours, and complex background interference remain key challenges in SAR target detection [19,20]. Despite these advances, common limitations remain: spatial enhancement still struggles to denoise without erasing weak structural cues; scale- and orientation-related information is under-exploited for irregular, scale-sensitive targets; and fixed model parameters adapt poorly to changing noise and clutter, limiting robustness and generalization.

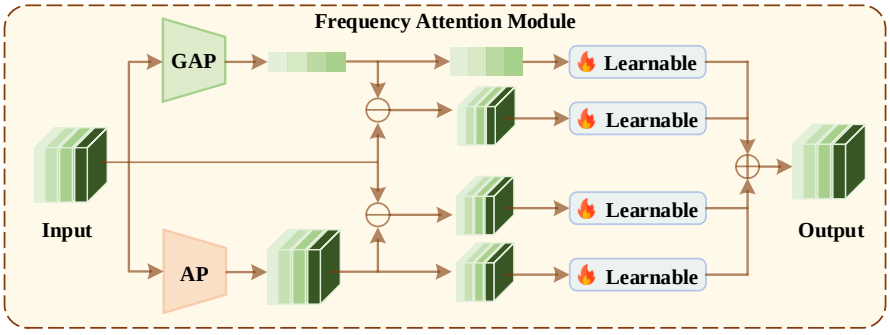


Figure 2. Existing frequency attention module.

We argue that overcoming these bottlenecks calls for explicitly introducing scale- and orientation-aware decomposition in the frequency domain. Such representations tend to concentrate structure while dispersing noise, and phase cues are less sensitive to magnitude fluctuations, providing complementary directional information. Moreover, because SAR targets are local and scale-dependent, coarse global frequency selection can suppress weak targets, motivating fine-grained sub-band modulation with input-adaptive refinement.

Based on these insights, we propose MSPaDet, a frequency-domain multi-scale phase-aware detection framework. MSPaDet introduces feature-level DTCWT to decompose features into multi-scale, multi-orientation complex sub-bands and modulates them using phase and energy cues for structure enhancement and speckle suppression. Within MSPaDet, SCFRDeno performs phase-consistency-guided sub-band reweighting and reconstruction to suppress coherent noise while preserving high-frequency details. The embedded PaSCA block then refines the fused features via phase-direction cues, spatial focusing, and region selection, improving stability under clutter and anisotropic scattering.

Our main contributions are as follows:

- We propose **MSPaDet**, a phase-aware detection framework using DTCWT for explicit directional sub-band modulation to enhance weak targets in speckle noise.

- We introduce **SCFRDeno–PaSCA**, which leverages phase consistency, local energy, and phase-direction cues for sub-band denoising and input-adaptive refinement.
- Experiments on MSAR, SAR-Aircraft-1.0, and SARDet-100K show superior accuracy and robustness with modest overhead, confirming practical utility.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the MSPaDet framework and its core modules, SCFRDeno and PaSCA. Section 4 describes the experimental setup and results, followed by ablation studies. Section 6 concludes this paper.

2. Related Work

2.1. Object Detection in SAR Imagery

Prior studies on SAR target detection can be broadly summarized into three directions: speckle suppression, detector design, and generalization across heterogeneous imaging conditions.

Speckle suppression. Coherent speckle degrades texture and edge cues in SAR imagery [21,22]. Existing speckle suppression methods can be broadly divided into imaging-level despeckling approaches and network-level feature enhancement methods [23–25]. Imaging-level methods, such as traditional filtering and model-based despeckling, are usually applied as preprocessing steps before detection. Although these methods can reduce speckle noise, overly aggressive despeckling may smooth weak target boundaries and erase local structural details that are important for detection. Recent works therefore tend to incorporate speckle suppression into detection networks, including soft-threshold quantization with contextual fusion [26], denoising-aware detection frameworks, and joint denoising–detection learning strategies such as DS-YOLO [27]. Compared with image-level preprocessing, network-level enhancement is generally more task-aware, but many existing methods still mainly rely on magnitude or spatial-domain modulation and rarely exploit phase consistency and directional scattering cues explicitly.

Detector design. From the perspective of framework design, existing SAR target detectors can be roughly categorized into two-stage and one-stage methods. Two-stage detectors first generate candidate proposals and then refine them for classification and localization, which can improve localization accuracy but usually introduce higher computational cost. One-stage detectors directly predict object categories and bounding boxes, and are widely adopted in SAR settings due to their favorable speed–accuracy trade-off [14,28–30]. Lightweight and anchor-free detectors further improve efficiency and flexibility for dense or small SAR targets [31,32]. Transformer-based and end-to-end detectors have also been introduced to enhance global context modeling and multi-scale representation for arbitrarily oriented targets [33,34], e.g., LD-Det [35]. Many methods also strengthen feature fusion and recalibration to improve small and irregular target detection [36]. However, most detector designs still focus on spatial-domain feature aggregation or architectural refinement, while explicit frequency-domain phase coherence and direction-aware scattering representations remain insufficiently explored.

Generalization and domain adaptation. Domain shifts caused by sensor and imaging variations have motivated adaptation-oriented studies, including domain-adaptive Transformer detectors for unlabeled multi-source SAR data [37] and unsupervised adaptation with cross-domain feature interaction and data balancing [38]. SARDet-100K further highlights the challenge of cross-scale generalization at the benchmark level [27]. Frequency-aware modeling has also been explored, such as polar-Fourier spatial–frequency fusion [39] and in-detector schemes that suppress speckle-dominant responses while retaining discriminative cues. Overall, most existing methods still focus on spatial-domain learning or coarse

frequency selection, with limited explicit modeling of phase coherence and directional scattering cues.

2.2. Adaptive Enhancement and Frequency-Domain Modulation

Spatially adaptive enhancement and region selection. Input-adaptive spatial enhancement has progressed beyond conventional re-weighting toward designs that emphasize receptive-field diversity and region-wise heterogeneity, as exemplified by RepLKNet [40], InternImage [41], and RFAConv [42].

Frequency-domain enhancement and selective modulation. Several works incorporate frequency information for selective enhancement, including frequency-guided strategies [43], frequency-domain attention for distillation [44], and frequency-enhanced pixel-attention frameworks [45]. However, most frequency-domain enhancement methods focus on magnitude modulation with global or static selection, and rarely model directional structure or phase consistency explicitly [46].

3. Method

3.1. Overview

The macro design philosophy of MSPaDet is to enhance SAR targets in a feature-level, frequency-domain, and phase-aware manner. Instead of relying on image-level despeckling, which may smooth weak target boundaries and high-frequency structural details, MSPaDet embeds speckle suppression directly into the detection pipeline. The design is guided by three principles: multi-scale and multi-directional decomposition for weak structure representation, phase-coherence-guided reweighting for speckle suppression, and adaptive refinement across spatial, channel, and regional dimensions for target-related response enhancement. Accordingly, SCFRDeno performs phase-aware frequency-domain reconstruction, while PaSCA further refines the reconstructed features in a task-adaptive manner.

The overall MSPaDet framework is shown in Figure 3. Built upon standard object detectors, MSPaDet introduces modular SCFRDeno modules for feature-level speckle suppression and structural enhancement. SCFRDeno performs explicit multi-scale wavelet decomposition and phase-aware sub-band reweighting, which suppresses coherent speckle while preserving high-frequency structural cues. The resulting representations are more consistent across scales and orientations, alleviating cross-scale inconsistency under anisotropic scattering and improving robustness to scale variation. With low parameter overhead and a plug-and-play design, SCFRDeno can be inserted into backbone stages or FPN fusion layers; in our implementation, it is placed at a key multi-scale fusion node to maximize its impact on feature quality and detection accuracy.

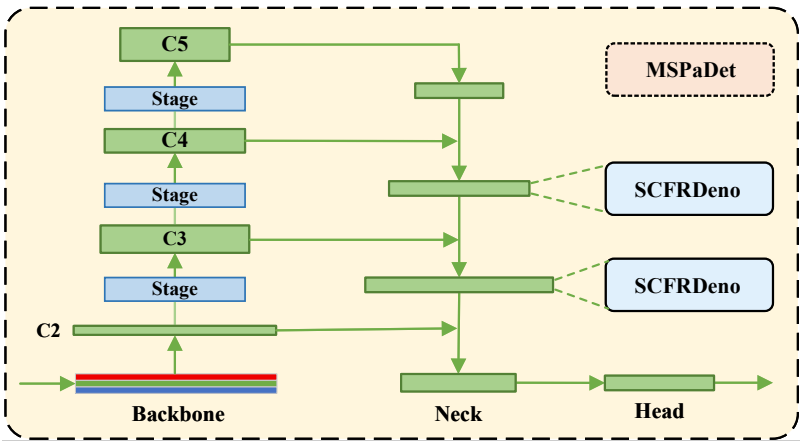


Figure 3. Simplified flowchart of the MSPaDet framework.

3.2. SCFRDeno

MSPaDet integrates the modular SCFRDeno module for feature-level speckle suppression and structural enhancement. Notably, PaSCA is an internal sub-module of SCFRDeno, which further performs input-adaptive refinement on the frequency-enhanced features. In this subsection, we introduce the frequency-domain formulation of SCFRDeno and define the intermediate feature F_{freq} that is subsequently refined by PaSCA. The detailed design of PaSCA is deferred to the dedicated section.

Algorithm 1 SCFRDeno

Input: Image tensor I
Output: Detection results $\mathcal{B}$
function INIT
  Initialize backbone f_{bb} (ResNet-50, outputs $\{C2, C3, C4, C5\}$)
  Initialize FPN base (out_channels $= 256$, start_level $= 1$, num_outs $= 5$)
  Initialize DTCWT (levels $= 2$) and PerBandNorm
  Initialize PATM with kernels (3×11) and (5×9)
  Initialize PaSCA blocks (Phase, Channel, Spatial, Group-Selection)
  Initialize detection head f_{head}
end function
function FORWARD(I)
  $X \leftarrow \text{Preprocess}(I)$
  $(C2, C3, C4, C5) \leftarrow f_{\text{bb}}(X)$
  Build FPN features $\{P_0, P_1, P_2, P_3, P_4\}$
  for $t = 0$ to 1 **do**
    $S_t \leftarrow P_t$
    $(L_t, \{H_t^{(d)}\}) \leftarrow \text{DTCWT}(S_t)$
    $(L_t, \{H_t^{(d)}\}) \leftarrow \text{PerBandNorm}(L_t, \{H_t^{(d)}\})$
    if Phase consistency enabled **then**
      $(L_t, \{H_t^{(d)}\}) \leftarrow \text{PhaseReweight}(L_t, \{H_t^{(d)}\})$
    end if
    $F_t^{\text{dtcwt}} \leftarrow \text{FuseBands}(L_t, \{H_t^{(d)}\})$
    $F_t^{\text{patm}} \leftarrow \text{PATM}(S_t)$
    $\widetilde{F}_t \leftarrow \text{Fuse}(F_t^{\text{dtcwt}}, F_t^{\text{patm}})$
    $P_t \leftarrow \text{PaSCA}(\widetilde{F}_t)$
  end for
  $\mathcal{B} \leftarrow f_{\text{head}}(\{P_0, P_1, P_2, P_3, P_4\})$
  return $\mathcal{B}$
end function

① **DTCWT decomposition.** SCFRDeno adopts Dual-Tree Complex Wavelet Transform (DTCWT) to obtain a multi-scale, multi-orientation complex representation. In the one-dimensional case, the low-pass and high-pass decompositions at the c-th layer are

$$a_c^{(t)}[k] = (x * h_0^{(t)}) \downarrow 2^j,$$

$$d_c^{(t)}[k] = (x * h_1^{(t)}) \downarrow 2^j,$$

$$t \in \{a, b\}.$$

The complex wavelet coefficient is formed by the real and imaginary outputs:

$$w_c = d_c^a + i d_c^b.$$

In the two-dimensional case, six directional sub-bands are obtained via separable row–column filtering, producing complex coefficients

$$W_{(c,\theta)} = D^{a}_{(c,\theta)} + iD^{b}_{(c,\theta)}, \qquad \theta \in \{\pm 15^\circ, \pm 45^\circ, \pm 75^\circ\}.$$

At each scale, DTCWT yields six oriented high-pass sub-bands along with a low-pass residual. The amplitude and phase are computed as

$$|W_{(c,\theta)}| = \sqrt{\left(D^{a}_{(c,\theta)}\right)^2 + \left(D^{b}_{(c,\theta)}\right)^2},$$

$$\angle W_{(c,\theta)} = \operatorname{atan2}\left(D^{b}_{(c,\theta)},\ D^{a}_{(c,\theta)}\right).$$

② **Phase consistency and sub-band reweighting.** We measure local phase coherence on each directional sub-band using

$$\rho_{c,\theta}(p) = \left| \frac{1}{|\mathcal{N}(p)|} \sum_{q \in \mathcal{N}(p)} \exp\left(i\left(\angle W_{c,\theta}(q) - \angle W_{c,\theta}(p)\right)\right) \right|,$$
$$\rho_{c,\theta}(p) \in [0, 1].$$

where $\mathcal{N}(p)$ is a spatial neighborhood. In practical applications, $\mathcal{N}(p)$ is defined as a $k \times k$ local window on each directional sub-band ($k = 3$ in this paper). We then apply phase-aware re-weighting to suppress low-coherence sub-band components:

$$\widetilde{W}_{(c,\theta)}(p) = \rho_{c,\theta}(p) \cdot W_{(c,\theta)}(p).$$

③ **Normalization and feature aggregation.** To prevent high-energy amplitude components from dominating subsequent feature learning while preserving phase information for structural orientation characterization, normalization is applied only to the amplitude:

$$\tilde{W}_{(c,\theta)}(p) = \frac{|\widetilde{W}_{(c,\theta)}(p)| - \mu_{(c,\theta)}}{\sigma_{(c,\theta)} + \varepsilon} e^{i\widetilde{W}_{(c,\theta)}(p)}.$$

The normalized coefficients preserve phase information for structural characterization. We concatenate the real and imaginary parts of all scales and orientations along the channel dimension to obtain a multi-scale, multi-directional complex representation. Here, "phase-aware" refers to the analytic sub-band phase of the DTCWT coefficients, which encodes local orientation and phase coherence, rather than the electromagnetic phase of raw complex SAR measurements.

The concatenated features are then compressed and aligned using 1×1 convolutions to produce F_{freq}. In MSPaDet, F_{freq} is refined by PaSCA and passed to the detection head for classification and localization.

3.3. PaSCA

Figure 4 illustrates the proposed PaSCA module, which refines the frequency-enhanced feature (e.g., F_{freq} from SCFRDeno) via complementary phase-, spatial-, and semantic-aware recalibration. PaSCA is composed of three components: (1) phase-aware component fusion, (2) cascaded channel–spatial adaptation, and (3) group-wise spatial gating.

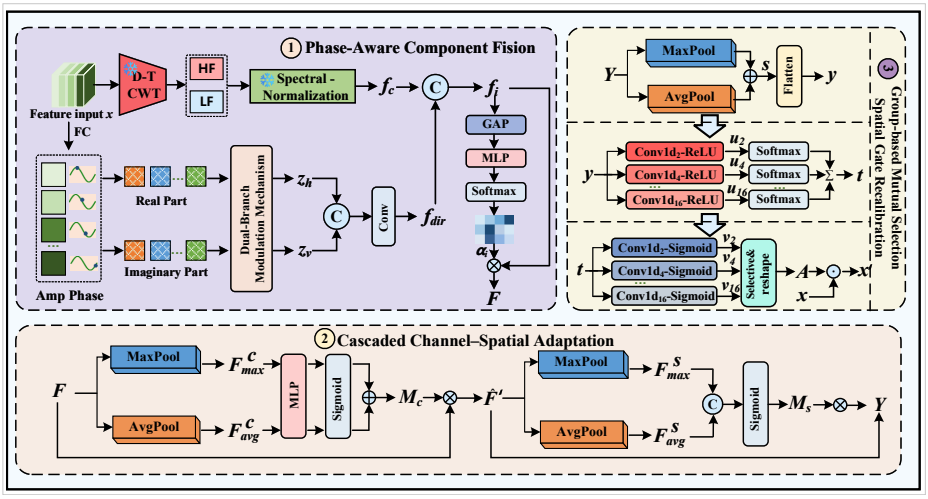


Figure 4. Diagram of the SCFRDeno and PaSCA modules.

① **Phase-aware component fusion.** To better accommodate anisotropic scattering and speckle interference, we model intermediate features with an amplitude–phase representation and perform direction-aware fusion. Given input features

$$\mathbf{x} \in \mathbb{R}^{B\times C\times H\times W},$$

mapped to the complex plane, each component is represented as a complex wave

$$\tilde{z}_c = |z_c| \odot e^{i\theta_c}, \qquad c = 1, 2, \ldots, n,$$

where $|z_c| \in \mathbb{R}^{B\times C\times H\times W}$ represents the characteristic amplitude, while $\theta_c \in \mathbb{R}^{B\times 1\times H\times W}$ denotes phase information used to describe structural orientation. The phase is generated through a lightweight per-channel mapping

$$\theta_c = \Theta(x_c, W^\theta).$$

where $x_c \in \mathbb{R}^{B\times C\times H\times W}$, the phase estimation module $\Theta(\cdot)$ outputs $\theta_C \in \mathbb{R}^{B\times C\times H\times W}$, and its learnable parameters satisfy $W^\theta \in \mathbb{R}^{1\times C\times k_h\times k_w}$. According to Euler's formula, the complex waveform can be decomposed into real and imaginary parts

$$\tilde{z}_c = |z_c| \odot \cos\theta_c + i\,|z_c| \odot \sin\theta_c.$$

To capture directional variability, we employ two directional branches

$$z_h = W_h^t(|z| \odot \cos\theta_h) + W_h^i(|z| \odot \sin\theta_h),$$
$$z_v = W_v^t(|z| \odot \cos\theta_v) + W_v^i(|z| \odot \sin\theta_v).$$

where $|z| \in \mathbb{R}^{B\times C\times H\times W}$ and $\theta_h, \theta_v \in \mathbb{R}^{B\times C\times H\times W}$. The learnable projection weights are defined as $W_h^t, W_h^i, W_v^t, W_v^i \in \mathbb{R}^{C_{\text{out}}\times C_{\text{in}}\times 1\times 1}$, and the modulated outputs satisfy $z_h, z_v \in \mathbb{R}^{B\times C_{\text{out}}\times H\times W}$. When two waveforms are superimposed, the resulting amplitude is influenced by their phase difference:

$$|z_r| = \sqrt{|z_i|^2 + |z_c|^2 + 2|z_i||z_c|\cos(\theta_c - \theta_i)}.$$

This motivates phase-modulated aggregation:

$$o_c = \sum_k \left(W_{ck}^t z_k \odot \cos\theta_k + W_{ck}^i z_k \odot \sin\theta_k\right).$$

In implementation, we concatenate the two directional outputs and obtain a compact direction-aware feature

$$f_{\text{dir}} = \text{Conv}(\text{Concat}[z_h, z_v]).$$

We further introduce a channel-semantic branch f_c and compute branch weights via global pooling and a two-layer MLP:

$$\alpha_i = \text{Softmax}(\text{MLP}(\text{GAP}(f_i))), \quad i \in \{h, v, c\}.$$

The final fused feature is obtained by adaptive aggregation:

$$F = \sum_{i \in \{h,v,c\}} \alpha_i f_i.$$

② **Cascaded channel–spatial adaptation.** We apply a lightweight two-stage correction on F to enhance salient targets and suppress background interference. First, channel statistics are computed by global average pooling and global max pooling:

$$F^c_{\text{avg}} = \text{GAP}(F), \qquad F^c_{\text{max}} = \text{GMP}(F).$$

They are fed into a shared two-layer MLP:

$$g(x) = W_1\, \delta(W_0 x),$$

yielding the channel attention map

$$M_c(\hat{F}) = \sigma\Big(g(F^c_{\text{avg}}) + g(F^c_{\text{max}})\Big) \quad \in \mathbb{R}^{B\times C\times 1\times 1}.$$

The channel-enhanced feature is

$$\hat{F}' = \hat{F} \odot (1 + \alpha M_c(\hat{F})).$$

Next, we compute channel-wise average and max maps

$$F^s_{\text{avg}} = \text{Avg}_c(\hat{F}'), \qquad F^s_{\text{max}} = \text{Max}_c(\hat{F}').$$

and obtain the spatial attention map

$$M_s(\hat{F}') = \sigma\Big(\mathcal{F}^{k_h\times k_w}\Big([F^s_{\text{avg}}; F^s_{\text{max}}]\Big)\Big).$$

Finally, spatial correction is applied as

$$Y = \hat{F}' \odot (1 + \beta M_s(\hat{F}')).$$

③ **Group-wise spatial gating.** To further accommodate heterogeneous clutter and spatially varying noise sensitivity, we introduce group-wise spatial gating for finer-grained region selection. Given an input feature map

$$x \in \mathbb{R}^{B\times C\times H\times W},$$

we first construct a compact spatial descriptor

$$s = \text{Avg}_c(x) + \text{Max}_c(x) \quad \in \mathbb{R}^{B\times H\times W},$$

$$y = \mathrm{vec}(s) \in \mathbb{R}^{B\times(HW)\times 1}.$$

We generate multi-granularity candidates with $G_1 = \{2, 4, 8, 16\}$ and compute grouping features

$$u_g = \mathrm{ReLU}(\mathrm{Conv1d}_g(y)), \qquad g \in G_1.$$

A micro-selective operator performs soft selection over candidates

$$\alpha_g = \frac{\exp(\phi(u_g))}{\sum_{g'\in G_1}\exp(\phi(u_{g'}))}, \qquad g \in G_1.$$

$$t = \sum_{g\in G_1} \alpha_g u_g.$$

We then construct another candidate set $G_1 = \{2, 4, 8, 16\}$ and generate gate candidates conditioned on t:

$$v_g = \sigma(\mathrm{Conv1d}_g(t)), \qquad g \in G_2.$$

The final gate vector is obtained by

$$a = \mathrm{Sel}(t, \{v_g \mid g \in G_2\}) \quad \in \mathbb{R}^{B\times(HW)\times 1}.$$

Finally, we reshape and apply the spatial gate:

$$A = \mathrm{reshape}(a) \in \mathbb{R}^{B\times 1\times H\times W},$$

$$x' = A \odot x.$$

The group configuration $G = \{2, 4, 8, 16\}$ is adopted to provide multi-granularity spatial gating. Smaller groups capture coarse and stable spatial responses, while larger groups enable finer regional selection for weak or small targets. This setting balances global stability and local adaptivity with a limited candidate space.

4. Experiments

4.1. Setup

4.1.1. Dataset

We evaluate MSPaDet on three public SAR detection benchmarks: SAR-Aircraft-1.0 (16,463 aircraft from seven classes), MSAR/HISEA-1 (60,396 targets from four classes), and SARDet-100K (16,598 images with 245,653 instances over six classes).

4.1.2. Implementation Details

All experiments were conducted on the autoDL platform, utilizing a single NVIDIA RTX 3090 GPU. Optimization was performed within the MMRotate framework, employing the DAdaptAdam optimizer with an automatically adjusted learning rate.

For MSAR, images were resized to 256 × 256 and randomly flipped ($p = 0.5$); training ran for 36 epochs with a batch size of 64, learning rate 1.0, and weight decay 0.05.

For SAR-Aircraft-1.0, raw images were cropped into 512 × 512 patches (200-pixel overlap) and trained for 12 epochs with a batch size of 32.

For SARDet-100K, images were resized to 512 × 512, flipped with $p = 0.5$, and trained for 12 epochs with a batch size of 12. The detailed information of each dataset is presented in Table 1. It should be noted that Pre. denotes the pretraining dataset; IN denotes the ImageNet-1K pretrained weights.

Table 1. Satellite and sensor details of the used datasets.

Dataset	Resolution	Band	Polarization	Satellites/Sensors
MSAR	≤1 m	C	HH/HV/VH/VV	HISEA-1
SAR-Aircraft-1.0	1 m	C	Single	GF-3
SARDet-100K	0.1–25 m	C/Ka/Ku/X	HH/HV/VH/VV/Single	Airborne SAR; GF-3; HISEA-1; RadarSat-2; S-1; TerraSAR-X; PALSAR-2

4.2. Test Results

4.2.1. Results on the MSAR Dataset

Table 2 compares MSPaDet with 20 state-of-the-art detectors on MSAR. MSPaDet achieves 69.53% mAP under AP'07 and 70.25% under AP'12, obtaining the best performance among all methods. Compared with DenoDet, MSPaDet improves mAP by about 1% while using fewer parameters and FLOPs, indicating a favorable accuracy–efficiency trade-off.

Table 2. Comparison with state-of-the-art methods on the MSAR dataset.

Method	Pre.	#P↓	FLOPs↓	AP'07					AP'12				
				mAP	Ship	Air	Brg	Oil	mAP	Ship	Air	Brg	Oil
One-stage													
FCOS [47]	IN	32.12 M	12.89 G	66.22	87.87	42.42	72.62	61.98	68.55	88.56	43.33	75.34	66.97
GFL [48]	IN	32.27 M	13.08 G	65.94	87.91	40.86	72.75	62.24	67.61	88.96	39.84	75.67	65.98
RepPoints [49]	IN	36.82 M	12.12 G	48.01	76.68	7.79	49.83	57.51	46.92	80.76	1.88	49.94	57.84
CenterNet [50]	IN	32.12 M	12.88 G	65.88	88.57	39.75	73.61	61.62	67.95	89.80	39.94	75.75	66.31
PVT-T [51]	IN	21.39 M	10.10 G	28.42	57.58	8.84	10.23	37.04	24.95	58.34	0.22	5.47	35.75
RetinaNet [52]	IN	36.39 M	13.14 G	32.41	62.73	1.42	22.17	45.54	32.02	63.67	1.66	19.78	45.27
TOOD [53]	IN	32.03 M	12.62 G	65.28	87.74	41.48	69.70	62.19	67.02	88.58	40.94	72.61	65.94
DDOD [54]	IN	32.27 M	29.02 G	59.50	85.29	61.86	70.12	20.74	61.01	87.45	65.39	73.49	17.73
YOLOF [55]	IN	42.41 M	6.58 G	45.09	68.09	7.79	57.51	46.95	43.29	70.42	1.01	58.77	44.95
YOLOX [56]	IN	8.94 M	2.13 G	67.10	88.59	62.19	60.16	57.44	69.12	89.31	65.76	63.23	58.19
S4 Det [57]	IN	32.72 M	12.09 G	64.01	86.64	37.50	69.73	62.19	65.26	87.91	35.44	71.82	65.86
Two-stage													
ConvNeXt [58]	IN	45.06 M	26.39 G	55.22	80.04	62.27	70.77	7.79	56.10	83.99	65.79	75.82	0.10
ConvNeXtV2 [59]	IN	105.00 M	40.71 G	55.08	80.06	62.29	70.18	7.79	56.00	84.16	65.55	73.56	0.70
LSKNet [60]	IN	30.98 M	22.52 G	56.83	80.01	62.31	77.21	7.79	57.44	85.56	65.75	77.42	1.04
BurgsVO [61]	IN	63.10 M	45.62 G	66.97	88.66	40.69	70.98	67.53	68.53	89.07	41.97	76.86	66.22
End2End													
DAB-DETR [62]	IN	43.70 M	10.49 G	42.58	78.95	2.17	50.90	38.31	42.87	82.90	0.28	51.06	37.26
Conditional DETR [63]	IN	43.35 M	9.79 G	40.96	77.33	8.50	52.83	25.17	38.71	81.56	1.30	53.04	20.34
DiffDet4 SAR [64]	IN	53.23 M	12.89 G	68.37	87.21	46.14	76.01	61.26	68.52	88.12	44.27	77.30	66.51
MSDFEN [65]	IN	40.23 M	13.41 G	68.38	87.43	45.23	77.03	63.81	68.73	88.84	45.37	78.50	66.96
DenoDet [66]	IN	34.23 M	12.89 G	68.60	88.10	46.57	77.62	62.11	69.91	89.44	44.77	78.30	67.11
MSPaDet (Ours)	IN	**27.25 M**	**7.75 G**	**69.53**	**89.35**	**48.6**	**76.92**	**63.25**	**70.25**	**90.05**	**49.3**	**77.32**	**64.45**

4.2.2. Results on the SAR-Aircraft-1.0 Dataset

As reported in Table 3, MSPaDet achieves 68.65% mAP and 69.78% AP^{12} on SAR-Aircraft-1.0, outperforming all 17 compared methods. Relative to S4Det, MSPaDet improves mAP by about 4% and AP^{12} by 4.6%, suggesting improved robustness to speckle and complex background structures.

4.2.3. Results on the SARDet-100K Dataset

Table 4 shows results on SARDet-100K. MSPaDet surpasses DiffDet4SAR by 0.9 mAP and maintains consistently stronger performance across diverse scenes. For small-target detection, MSPaDet further outperforms the strongest competing model, MSDFEN, by 0.3 percentage points in mAP_s, providing direct evidence of its advantage on small objects. It also exhibits smaller accuracy drops across target types and imaging settings, reflecting better robustness.

Table 3. Comparison with state-of-the-art methods on the SAR-Aircraft-1.0 dataset.

Method	Pre.	FLOPs	Average Precision (AP'07)								Average Precision (AP'12)							
			mAP	A220	A320	A330	ARJ21	B737	B787	Other	mAP	A220	A320	A330	ARJ21	B737	B787	Other
One-stage																		
FCOS [47]	IN	51.58 G	57.13	56.77	87.69	58.15	57.60	34.56	46.87	58.29	58.26	59.13	89.66	59.11	57.52	33.86	47.55	61.00
GFL [48]	IN	32.27 G	61.40	54.40	81.84	87.38	57.03	40.66	56.32	52.14	62.94	55.69	85.11	89.25	60.60	40.48	56.61	52.88
RepPoints [49]	IN	48.50 G	61.63	58.77	84.25	80.78	55.78	36.00	54.23	61.60	62.59	60.22	87.53	81.41	57.38	34.51	54.30	62.79
TOOD [53]	IN	50.53 G	57.16	48.44	76.70	86.36	55.01	29.86	48.31	55.45	57.48	48.31	78.41	88.49	55.01	28.41	48.00	55.71
DDOD [54]	IN	45.60 G	57.16	47.52	81.84	90.62	54.45	29.81	39.04	56.89	57.60	46.94	85.95	90.81	55.90	29.33	36.90	57.37
YOLOF [55]	IN	26.33 G	60.75	54.74	85.25	83.31	55.50	30.80	54.14	61.49	62.21	55.47	89.35	85.81	57.87	28.85	54.92	63.21
YOLOX [56]	IN	8.53 G	58.15	58.43	84.11	92.04	46.51	18.69	51.13	56.17	59.55	59.41	85.77	92.38	50.01	17.38	55.15	56.76
S4 Det [57]	IN	48.39 G	64.67	59.44	89.41	89.75	59.79	41.91	50.96	57.43	65.15	62.50	90.07	90.64	59.26	39.13	55.26	58.93
Two-stage																		
PRDet [67]	IN	51.00 G	65.37	52.37	90.82	92.01	63.15	42.77	58.89	57.59	66.35	53.26	91.08	92.23	66.16	43.71	59.60	58.38
ConvNeXt [58]	IN	63.85 G	61.91	57.94	86.75	92.03	63.85	33.08	43.56	56.14	61.48	58.01	88.21	92.16	62.11	30.67	43.56	55.61
ConvNeXt V2 [59]	IN	0.12 T	62.54	55.63	88.27	91.66	54.23	30.78	56.88	60.32	63.22	57.21	89.38	91.80	54.43	28.74	57.69	63.26
LSKNet [60]	IN	53.73 G	62.08	53.65	93.56	91.33	53.35	31.84	51.34	59.48	62.76	54.32	93.71	91.71	55.30	30.06	51.97	62.29
End2End																		
DAB-DETR [62]	IN	28.94 G	48.12	54.32	82.89	13.34	56.76	29.02	48.84	51.65	48.66	55.30	85.60	13.47	58.11	27.70	48.54	51.92
Conditional DETR [63]	IN	28.09 G	56.75	48.24	83.88	73.33	56.25	31.94	55.95	47.69	57.52	48.77	87.80	74.97	56.57	30.42	56.49	47.64
DiffDet4 SAR [64]	IN	50.49 G	67.46	63.22	91.14	90.87	56.08	45.32	60.24	65.35	69.06	66.03	91.93	92.34	57.30	47.35	61.24	67.21
MSDFEN [65]	IN	73.27 G	67.22	64.32	88.76	90.65	55.21	44.82	60.84	65.97	69.34	66.03	92.93	92.34	57.30	46.35	62.24	68.21
DenoDet [66]	IN	48.53 G	63.10	59.32	85.04	88.32	54.08	40.82	50.54	63.55	64.06	61.03	87.93	89.34	55.30	39.35	50.24	65.21
MSPaDet (Ours)	IN	**68.38 G**	**68.65**	**65.73**	**91.49**	**92.04**	**56.34**	**45.40**	**62.33**	**67.22**	**69.78**	**68.01**	**92.54**	**92.34**	**57.29**	**46.51**	**63.58**	**68.19**

Table 4. Comparison with state-of-the-art methods on the SARDet-100K dataset.

Method	Pre.	FLOPs	#Params↓	mAP	AP@50	AP@75	APs	APm	APl	Ship	Aircraft	Car	Tank	Bridge	Harbor
One-stage															
FCOS [47]	IN	51.57 G	32.13 M	52.52	85.82	54.93	47.01	66.13	57.82	59.79	55.44	60.75	41.78	34.17	63.44
GFL [48]	IN	52.36 G	32.27 M	54.71	84.86	58.57	49.14	66.99	60.15	63.62	57.33	61.99	44.50	36.11	64.74
RepPoints [49]	IN	48.49 G	36.82 M	51.36	86.13	53.69	46.36	62.96	53.48	60.55	55.20	60.83	40.39	34.82	56.41
ATSS [68]	IN	51.57 G	32.13 M	54.65	87.30	57.95	49.59	67.64	58.67	61.23	55.64	61.47	45.90	36.92	67.18
PVT-T [51]	IN	42.19 G	21.43 M	45.80	77.25	48.70	37.71	59.23	53.05	53.00	52.61	58.73	29.90	22.21	58.81

Table 4. *Cont.*

Method	Pre.	FLOPs	#Params↓	mAP	AP@50	AP@75	APs	APm	APl	Ship	Aircraft	Car	Tank	Bridge	Harbor
TOOD [53]	IN	50.52 G	30.03 M	54.35	86.58	58.11	49.90	66.42	58.30	61.98	55.31	62.23	45.66	36.34	64.94
DDOD [54]	IN	45.58 G	32.21 M	53.72	86.34	56.93	49.03	64.40	57.72	62.09	55.78	62.18	43.68	36.04	62.57
YOLOF [55]	IN	26.32 G	42.46 M	42.53	74.65	42.88	33.43	55.89	53.27	52.32	52.34	52.41	22.56	23.44	52.12
YOLOX [56]	IN	8.53 G	8.94 M	33.78	66.47	31.01	28.19	42.76	28.65	45.78	46.53	53.13	25.96	12.84	18.65
S4 Det [57]	IN	48.38 G	32.72 M	55.71	87.02	59.02	50.07	68.09	60.69	64.84	58.54	64.67	44.78	36.81	64.98
Two-stage															
ConvNeXt [58]	IN	63.84 G	45.07 M	52.85	85.22	56.98	45.37	64.25	58.31	60.25	57.05	61.83	37.82	36.51	63.65
ConvNeXtV2 [59]	IN	0.12 T	0.11 G	53.61	85.71	58.60	47.33	64.37	59.27	61.18	55.53	62.93	39.35	38.86	63.79
BurgsVO [61]	IN	63.97 G	45.20 M	53.98	85.35	57.98	47.63	66.38	60.44	61.38	58.31	62.83	38.82	37.90	64.65
End2End															
DAB-DETR [62]	IN	28.94 G	43.70 M	43.01	77.84	42.80	34.52	56.04	52.32	52.86	50.02	49.17	23.76	28.17	54.77
DiffDet4 SAR [64]	IN	69.82 G	54.74 M	55.79	85.51	59.46	50.33	67.40	60.30	64.81	57.06	63.79	45.56	36.22	66.94
MSDFEN [65]	IN	82.73 G	75.78 M	56.12	86.21	59.87	50.51	68.38	61.22	65.21	58.36	64.16	46.23	37.73	65.07
DenoDet [66]	IN	52.69 G	65.78 M	55.32	85.51	59.25	50.33	67.40	60.30	64.61	57.06	63.36	45.49	36.09	66.87
MSPaDet (Ours)	IN	**64.60 G**	**49.50 M**	**56.69**	**86.32**	**61.17**	**50.80**	**68.30**	**62.50**	**65.94**	**58.33**	**64.32**	**45.13**	**37.81**	**68.50**

4.2.4. Cross-Dataset Transfer Evaluation

To further assess the transferability of MSPaDet beyond independent in-domain evaluation, we conduct a cross-dataset aircraft-level transfer experiment. Specifically, the model is trained on SAR-Aircraft-1.0 and directly evaluated on the aircraft subset of SARDet-100K without target-domain fine-tuning. Since SAR-Aircraft-1.0 contains seven aircraft subclasses, these subclasses are merged into a single aircraft category for aircraft-level evaluation.

As shown in Table 5, MSPaDet achieves an AP of 0.4101 and a Recall of 0.6100 under the cross-dataset setting. For contextual reference, the in-domain mAP on SAR-Aircraft-1.0 and SARDet-100K is 0.6865 and 0.5833, respectively. These results suggest that the proposed method maintains non-trivial transfer capability under heterogeneous imaging conditions, even without adaptation to the target domain.

Table 5. In-domain and cross-dataset aircraft-level evaluation results.

Category	Metric	Score
In-domain reference		
SAR-Aircraft-1.0	mAP	0.6865
SARDet-100K	mAP	0.5833
Cross-dataset transfer		
SAR-Aircraft-1.0 → SARDet-100K (aircraft subset)	AP	0.4101
SAR-Aircraft-1.0 → SARDet-100K (aircraft subset)	Recall	0.6100

We note that this experiment is intended to provide supplementary evidence of cross-dataset transferability rather than a strict leave-one-sensor-out evaluation, since reliable image-level sensor labels are not consistently available for all samples in the target dataset.

4.2.5. Comparison with Traditional Despeckling-Based Detection Pipelines

To further verify the necessity of the proposed phase-aware feature-level denoising strategy, we compare MSPaDet with traditional image-level despeckling methods embedded into the same detection pipeline. Specifically, Lee filtering, non-local means (NLM), and SAR-BM3D are applied to both the training and testing images as preprocessing modules, while the annotation files remain unchanged. The same baseline detector is then retrained and evaluated on the corresponding despeckled datasets.

As shown in Figure 5, traditional despeckling methods improve the baseline detector to some extent, but their gains remain limited. MSPaDet achieves the highest mAP, outperforming Lee, NLM, and SAR-BM3D by 5.38, 5.10, and 4.57 percentage points, respectively. These results indicate that image-level despeckling can suppress speckle interference, but may also weaken weak target boundaries and local structural details. In contrast, the proposed feature-level phase-aware denoising better balances speckle suppression and structure preservation for SAR target detection.

4.3. Ablation Study

4.3.1. Ablation on Frequency-Domain Modeling

We investigate whether jointly modeling energy and phase in the frequency domain is more effective than using either cue alone. As reported in Table 6, we ablate SCFRDeno by retaining only the energy branch, only the phase branch, or the full energy–phase design. Using only energy improves mAP on MSAR by 1.12%, and using only phase yields a 1.36% gain, while the full dual-branch collaboration achieves a 3.00% improvement over the baseline without additional computational overhead. These results suggest clear complementarity between energy and phase cues for capturing structural details and

directional patterns under speckle contamination. In addition, FP64 provides only a marginal 0.26% mAP increase but introduces substantial overhead; we therefore use FP32 in all experiments to balance accuracy and efficiency.

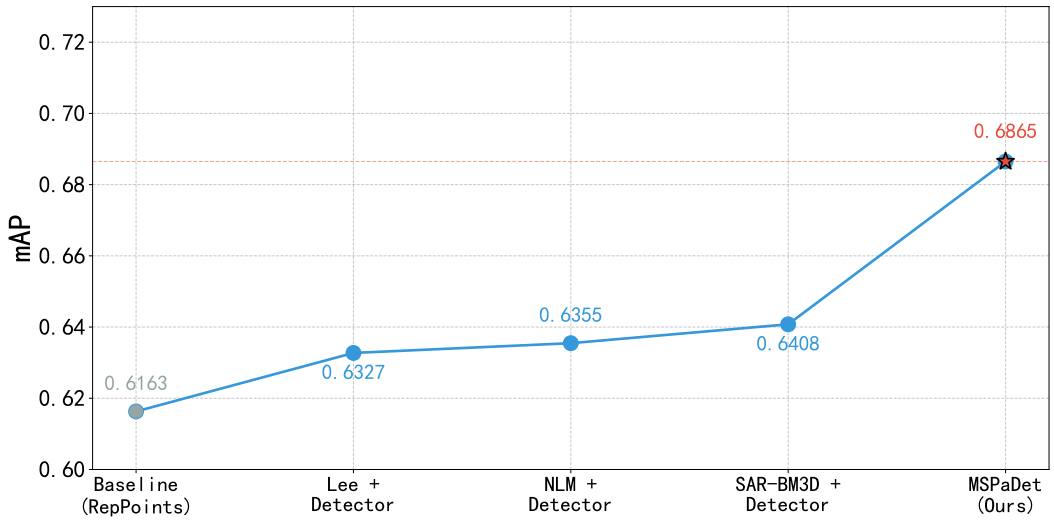


Figure 5. Comparison with traditional despeckling methods embedded in the same detection pipeline on SAR-Aircraft-1.0.

Table 6. Ablation on frequency-domain decomposition.

DTCWT/Norm.	Phase Reweight	FP	MSAR			SAR-Aircraft-1.0			SARDet-100K	
			FPS	mAP(07)	mAP(12)	FPS	mAP(07)	mAP(12)	FPS	mAP
×	×	32	230.4	65.52	66.85	167.6	61.63	62.76	101.3	53.21
×	✓	32	230.1	66.88	67.09	167.4	63.01	63.40	102.4	54.91
✓	×	32	225.0	66.64	68.26	167.3	59.72	60.49	101.4	53.81
✓	×	64	220.3	66.71	68.57	152.3	60.58	61.05	101.3	54.32
✓	✓	32	215.0	68.52	70.31	166.4	67.64	68.61	101.9	55.68

4.3.2. Ablation Study of Dual-Branch Modeling

Table 7 compares the spatial-only baseline with a variant that introduces DTCWT-based frequency-domain decomposition. Incorporating DTCWT yields consistent improvements across metrics, including a 1.34% mAP gain on MSAR, with the largest benefit on small targets. These results support the advantage of frequency-domain modeling for speckle-robust and direction-sensitive feature representation. Notably, SCFRDeno is a SAR-oriented phase-aware enhancement built on DTCWT sub-bands, rather than a plain stack of wavelet filters, enabling adaptive suppression of speckle-dominant components while preserving structural details.

Table 7. Ablation on dual-branch modeling.

DTCWT/Norm.	PaSCA	MSAR		SAR-Aircraft-1.0		SARDet-100K
		mAP(07)	mAP(12)	mAP(07)	mAP(12)	mAP
×	×	66.52	67.85	61.63	62.76	51.66
✓	×	67.86	68.81	64.20	64.84	54.30
✓	✓	69.53	70.25	68.65	69.78	56.69

4.3.3. Sensitivity Analysis of Key Hyperparameters

To justify the default phase-coherence window size, we conduct a sensitivity analysis by evaluating $k = \{1, 3, 5, 7\}$ on MSAR, SAR-Aircraft-1.0, and SARDet-100K while keeping

all other settings unchanged. As shown in Figure 6, the performance on all three datasets peaks at $k = 3$, indicating that this window size provides the best overall trade-off between local stability and preservation of weak structural details. When $k = 1$, the neighborhood is too small to support stable local phase-coherence estimation and is therefore more sensitive to speckle fluctuations. In contrast, larger windows such as $k = 5$ and $k = 7$ tend to mix target and background responses, which weakens the local structural details of weak targets. Therefore, $k = 3$ is adopted as the default setting.

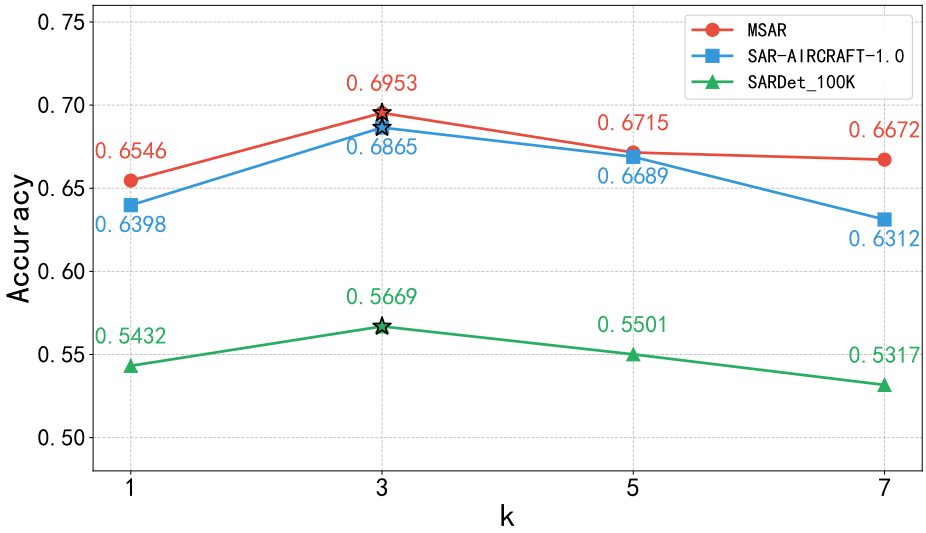


Figure 6. Effect of phase-coherence window size k on detection performance.

We further analyze the effect of DTCWT decomposition depth by comparing $L = \{1, 2, 3, 4\}$ on the three datasets. As shown in Figure 7, $L = 2$ yields the best overall performance across all three datasets. A shallow decomposition such as $L = 1$ is insufficient to capture multi-scale structural responses, whereas deeper decompositions such as $L = 3$ and $L = 4$ introduce additional redundancy and may dilute weak local target details. Therefore, a two-level DTCWT decomposition is used as the default setting.

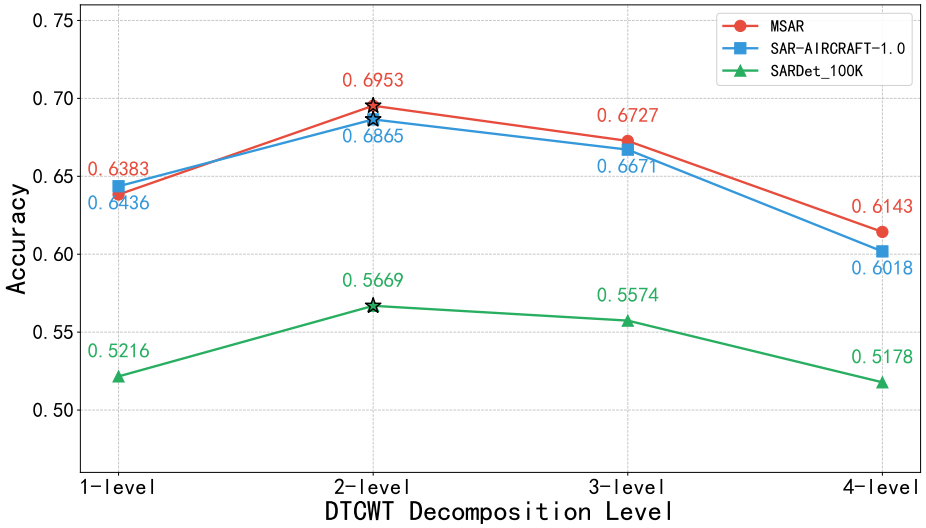


Figure 7. Effect of DTCWT decomposition depth on detection performance.

4.3.4. Complexity Analysis

To provide a clearer analysis of the computational overhead introduced by the proposed modules, we compare the parameter counts and FLOPs of four variants: the baseline

detector, baseline + PaSCA, baseline + SCFRDeno, and the full MSPaDet. As shown in Figure 8, the baseline detector contains 28.78 M parameters and 55.07 G FLOPs. After adding PaSCA, the model reaches 33.59 M parameters and 59.82 G FLOPs, corresponding to increases of 16.7% and 8.6%, respectively. After adding SCFRDeno, the model reaches 44.06 M parameters and 61.85 G FLOPs, corresponding to increases of 53.1% and 12.3%, respectively. The full MSPaDet contains 49.50 M parameters and 64.60 G FLOPs, corresponding to overall increases of 72.0% and 17.3% over the baseline detector. These results show that PaSCA is relatively compact, while SCFRDeno introduces a larger parameter overhead but keeps the FLOP increase moderate. Overall, MSPaDet achieves improved detection performance with moderate computational overhead.

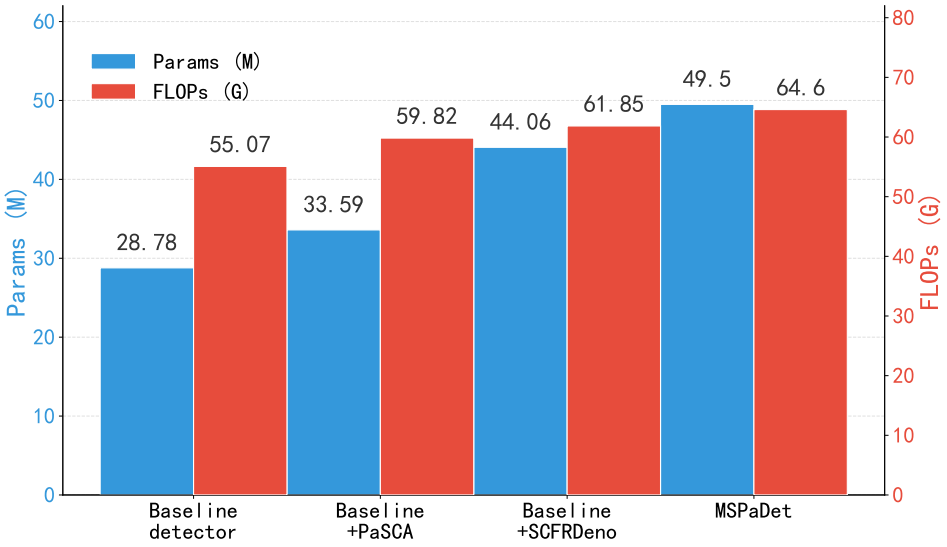


Figure 8. Parameter and FLOPs comparison of MSPaDet variants relative to the baseline detector.

4.4. Visualization of Detection Results

To further support the above findings, we visualize feature responses with Grad-CAM. On MSAR, the baseline FCOS exhibits pronounced high-frequency speckle artifacts in FPN-level features, whereas MSPaDet produces substantially cleaner and more target-focused activations. On SAR-Aircraft-1.0, MSPaDet better separates targets from cluttered backgrounds and suppresses spurious responses around highly reflective structures. Overall, the visual evidence is consistent with the quantitative gains, indicating that SCFRDeno enhances structural cues while reducing speckle-dominant interference in the feature space.

4.4.1. Small-Target Analysis and Feature Visualization

To further explain the small-target improvement, we visualize representative tiny targets smaller than 5×5 pixels from MSAR, SAR-Aircraft-1.0, and SARDet-100K. As shown in Figure 9, MSPaDet produces more concentrated feature responses than MSDFEN around tiny targets and better suppresses background speckle and clutter, indicating improved target–background separability. This benefit mainly comes from the multi-scale and multi-directional representation of DTCWT, the suppression of random speckle by phase-coherence modeling, and the subsequent feature refinement that enhances weak target responses and boundary cues.

4.4.2. False Positive Cases

As shown in Figure 10, MSPaDet reduces false alarms compared with the FCOS baseline. In MSAR, FCOS is prone to triggering on cluttered high-energy background regions (e.g., ports and coastlines), whereas MSPaDet suppresses these responses after

frequency-domain reconstruction (Figure 10a). On SAR-Aircraft-1.0, MSPaDet better separates structural high-frequency cues from random noise, leading to fewer spurious detections (Figure 10b).

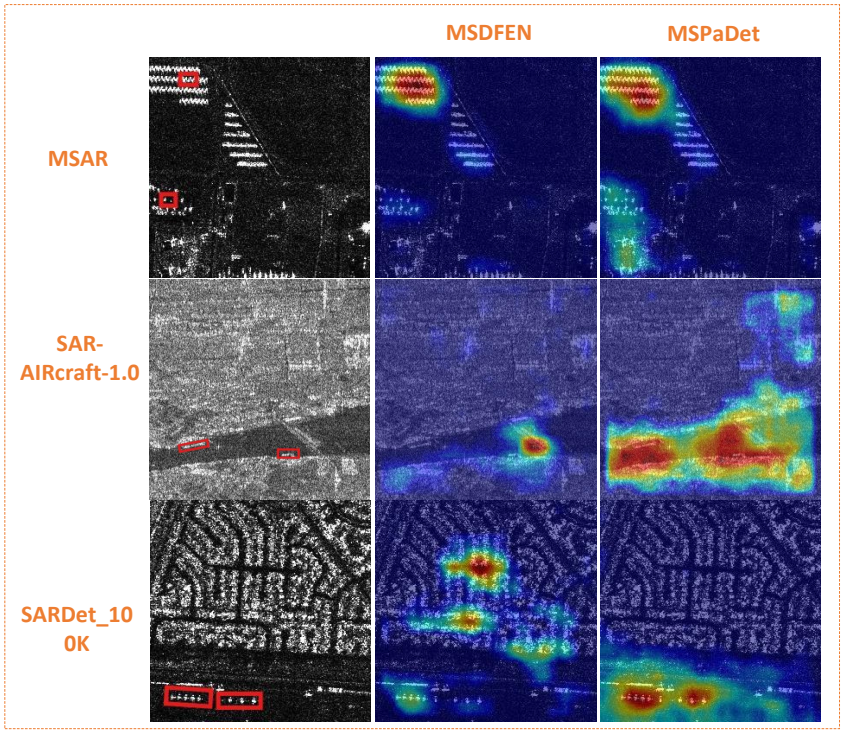


Figure 9. Qualitative comparison and feature-map visualization for tiny targets within 5 × 5 pixels across MSAR, SAR-Aircraft-1.0, and SARDet-100K.

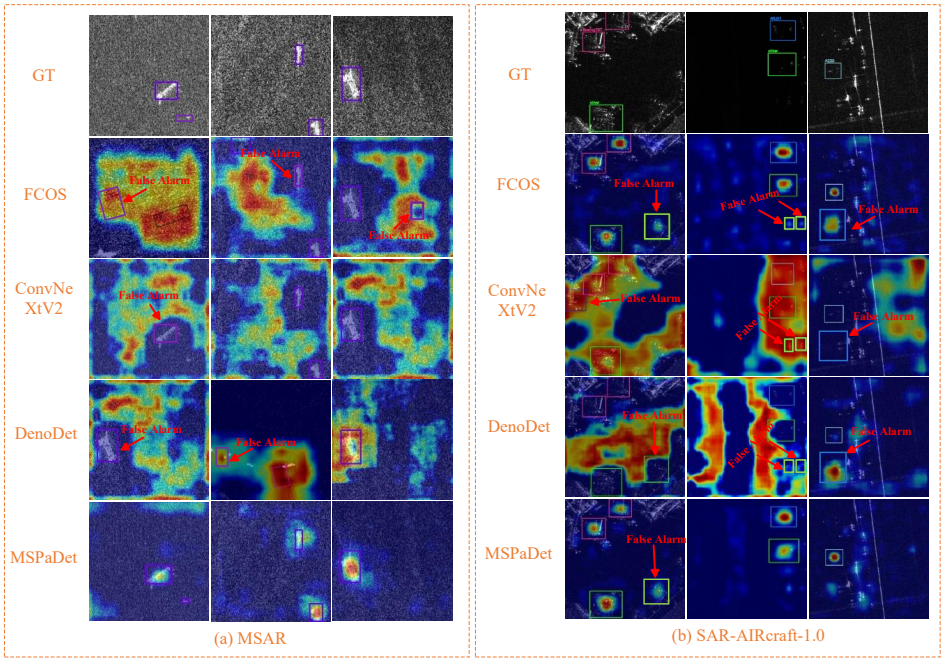


Figure 10. Comparison of feature-map denoising produced by different detection methods.

4.4.3. Miss Detection Cases

Figure 11 shows that MSPaDet alleviates missed detections for small or low-reflectivity targets. On MSAR, SCFRDeno enhances weak yet structurally consistent signals via subband reweighting, producing more concentrated activations on micro targets and improv-

ing recall. On SAR-Aircraft-1.0, the embedded PaSCA refinement further strengthens salient responses in cluttered scenes, helping maintain high recall.

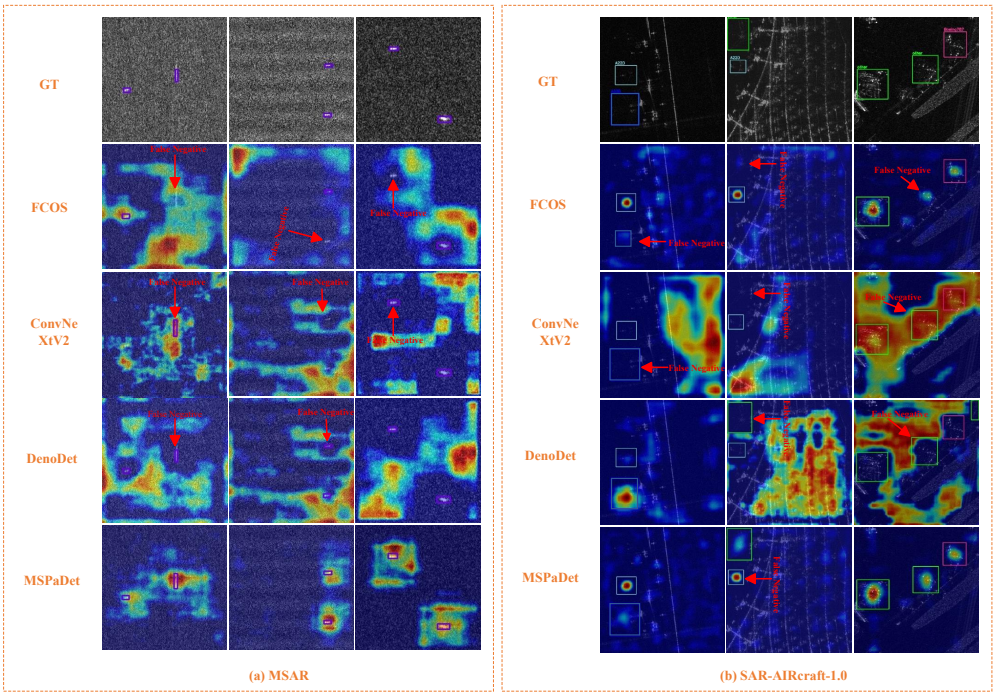


Figure 11. Misdetection examples for SAR targets; each subfigure shows one case.

4.4.4. Misclassification

As illustrated in Figure 12, MSPaDet yields more class-discriminative Grad-CAM responses in complex scenes. The combination of frequency-domain enhancement and spatial refinement produces more target-focused activations and suppresses background-induced ambiguity, which helps reduce misclassification.

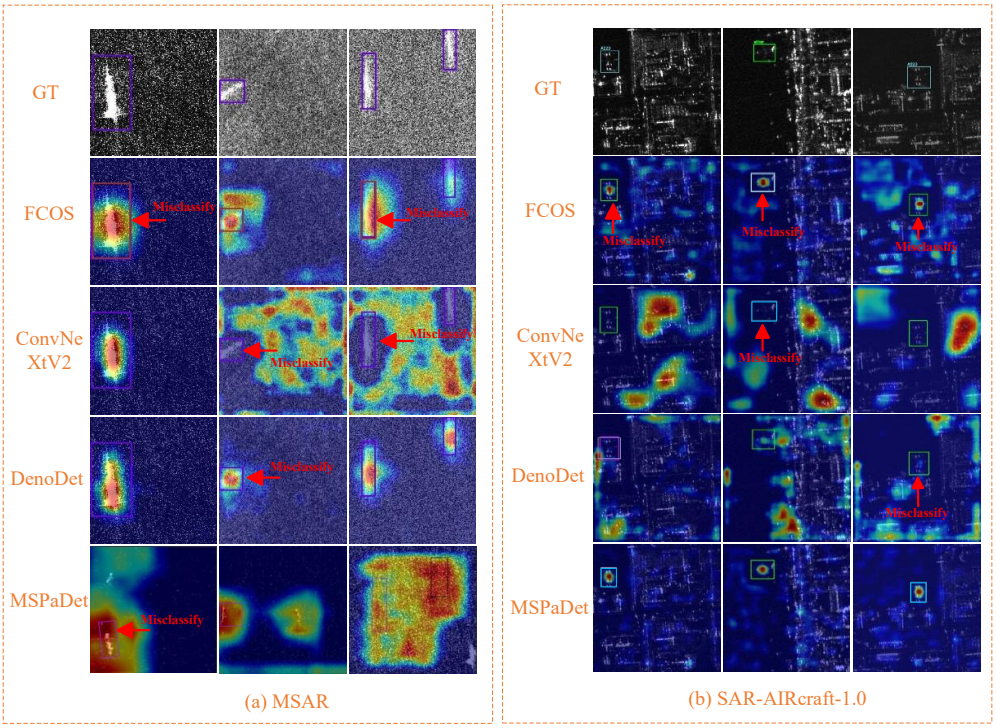


Figure 12. Misclassification examples for SAR targets; each subfigure shows one case.

5. Discussion

This work addresses the coexistence of weak target responses and speckle in high-frequency SAR features. MSPaDet distinguishes them by exploiting directional phase coherence, preserving stable scattering cues while suppressing unstable high-frequency interference. Experiments on MSAR, SAR-Aircraft-1.0, and SARDet-100K show consistent improvements, especially on SAR-Aircraft-1.0, where aircraft targets often appear as sparse and incomplete scattering structures. Ablation studies confirm that energy and phase provide complementary cues: energy highlights strong scatterers, while phase coherence captures stable directional structures.

Compared with Lee filtering, NLM, and SAR-BM3D, MSPaDet better matches the detection objective by performing denoising in feature space rather than smoothing the input image. It also complements existing SAR detectors by adding an explicit scattering-coherence constraint.

However, feature denoising alone cannot fully address cross-dataset domain shifts. In addition, the phase information comes from DTCWT feature coefficients rather than raw complex SAR data, and SCFRDeno increases model parameters. Future work will explore raw complex phase information, phase-aware adaptation, and lightweight sub-band selection.

6. Conclusions

We present MSPaDet for SAR target detection, which leverages complex wavelet sub-bands and phase-coherence modulation for speckle-robust feature enhancement. MSPaDet achieves consistent state-of-the-art performance on MSAR, SAR-Aircraft-1.0, and SARDet-100K, with a 1.0% mAP improvement in MSAR and a 0.57% gain over SARDet-100K.

Author Contributions: N.C.: Conducted the experiments, processed and analyzed the data, implemented the software, and drafted the manuscript. X.X.: Conceived and oversaw the study; provided expert guidance throughout the research. Y.L.: Wrote and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: Science Research Projects of Hunan Provincial Education Department (Grant No. 25B0310).

Data Availability Statement: The data analyzed in this study are publicly available from the MSAR dataset (https://radars.ac.cn/web/data/getData?dataType=MSAR (accessed on 20 September 2025), the SAR-Aircraft-1.0 dataset (https://radars.ac.cn/web/data/getData?newsColumnId=f896637b-af23-4209-8bcc-9320fceaba19, accessed on 20 September 2025), and the SARDet-100K dataset (https://github.com/zcablii/SARDet_100K, accessed on 20 September 2025). The code used in this study is available from the corresponding author upon reasonable request.

Acknowledgments: Acknowledgments:This work was supported in part by the Natural Science Foundation of Hunan Province under Grant 2025JJ50397, and by the Science Research Projects of Hunan Provincial Education Department under Grant No. 25B0310. We gratefully acknowledge the financial support from both funding agencies. We also extend our sincere thanks to the following individuals for their specific contributions:Xuyu Xiang: Conceived and oversaw the study; provided expert guidance throughout the research.Yuanjing Luo: Wrote and revised the manuscript.

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

References

1. Tirandaz, Z.; Akbarizadeh, G.; Kaabi, H. PolSAR Image Segmentation Based on Feature Extraction and Data Compression Using Weighted Neighborhood Filter Bank and Hidden Markov Random Field–Expectation Maximization. *Measurement* **2020**, *153*, 107432. [CrossRef]
2. Quan, S.; Zhang, T.; Xing, S.; Wang, X.; Yu, Q. Maritime Ship Detection with Concise Polarimetric Characterization Pattern. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *131*, 103954. [CrossRef]

3. Chen, S.; Cui, X.; Wang, X.; Xiao, S. Speckle-Free SAR Image Ship Detection. *IEEE Trans. Image Process.* **2021**, *30*, 5969–5983. [CrossRef]
4. Yasir, M.; Liu, S.; Xu, M.; Wan, J.; Sheng, H.; Nazir, S.; Zhang, X.; Isiacik Colak, A.T. YOLOv8-BYTE: Ship Tracking Algorithm Using Short-Time Sequence SAR Images for Disaster Response Leveraging GeoAI. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *128*, 103771. [CrossRef]
5. Karwowska, K.; Slesinski, J.; Wierzbicki, D. Effectiveness of YOLO Variants for Small Object Detection in SAR Images Using a New Dataset. *Sci. Rep.* **2025**, *15*, 45405. [CrossRef]
6. Liu, W.; Qin, J. Focus and Learn: Boosting Deep Multi-View Clustering via Hard Instance Awareness. *Inf. Fusion* **2026**, *127*, 103724. [CrossRef]
7. Wu, B.; Liu, C.; Chen, J. A Review of Spaceborne High-Resolution Spotlight/Sliding Spotlight Mode SAR Imaging. *Remote Sens.* **2025**, *17*, 38. [CrossRef]
8. Wang, Y.; Li, J.; Yang, J.; Sun, B. A Novel Spaceborne Sliding Spotlight Range Sweep Synthetic Aperture Radar: System and Imaging. *Remote Sens.* **2017**, *9*, 783. [CrossRef]
9. Zhang, Z.; Yu, W.; Zheng, M.; Zhao, L.; Zhou, Z.X. Phase Mismatch Calibration for Dual-Channel Sliding Spotlight SAR-GMTI. *Remote Sens.* **2022**, *14*, 617. [CrossRef]
10. Liu, T.; Yang, Z.; Marino, A.; Gao, G.; Yang, J. Robust CFAR Detector Based on Truncated Statistics for Polarimetric Synthetic Aperture Radar. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 6731–6747. [CrossRef]
11. Zalpour, M.; Akbarizadeh, G.; AlaeiSheini, N. A New Approach for Oil Tank Detection Using Deep Learning Features with Control False Alarm Rate in High-Resolution Satellite Imagery. *Int. J. Remote Sens.* **2020**, *41*, 2239–2262. [CrossRef]
12. Tang, G.; Zhao, H.; Claramunt, C.; Zhu, W.; Wang, S.; Wang, Y.; Ding, Y. PPA-Net: Pyramid Pooling Attention Network for Multi-Scale Ship Detection in SAR Images. *IEEE Trans. Pattern Anal. Mach. Intelling* **2023**, *15*, 2855. [CrossRef]
13. 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]
14. 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]
15. Li, Y.; Li, X.; Li, W.; Hou, Q.; Liu, L.; Cheng, M.; Yang, J. SARDet-100K: Towards Open-Source Benchmark and ToolKit for Large-Scale SAR Object Detection. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Vancouver, BC, Canada, 10–15 December 2024.
16. Zhang, X.; Yang, X.; Li, Y.; Yang, J.; Cheng, M.; Li, X. RSAR: Restricted State Angle Resolver and Rotated SAR Benchmark. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 10–17 June 2025.
17. Du, Y.; Chen, Y.; Huang, L.; Yang, Y.; Ghamisi, P.; Du, Q. SUMMIT: A SAR Foundation Model with Multiple Auxiliary Tasks Enhanced Intrinsic Characteristics. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *141*, 104624. [CrossRef]
18. Ma, W.; Yang, X.; Zhu, H.; Wang, X.; Yi, X.; Wu, Y.; Hou, B.; Jiao, L. NRENet: Neighborhood Removal-and-Emphasis Network for Ship Detection in SAR Images. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *131*, 103927. [CrossRef]
19. Zhang, T.; Zhang, X.; Ke, X.; Zhan, X.; Shi, J.; Wei, S.; Pan, D.; Li, J.; Su, H.; Zhou, Y.; et al. LS-SSDD-v1.0: A Deep Learning Dataset Dedicated to Small Ship Detection from Large-Scale Sentinel-1 SAR Images. *Remote Sens.* **2020**, *12*, 2997. [CrossRef]
20. Ke, X.; Zhang, T.; Shao, Z. Scale-aware dimension-wise attention network for small ship instance segmentation in synthetic aperture radar images. *J. Appl. Remote Sens.* **2023**, *17*, 046504. [CrossRef]
21. Lee, J.S. Speckle Analysis and Smoothing of Synthetic Aperture Radar Images. *Comput. Graph. Image Process.* **1981**, *17*, 24–32. [CrossRef]
22. Parrilli, S.; Poderico, M.; Angelino, C.V.; Verdoliva, L. A Nonlocal SAR Image Denoising Algorithm Based on LLMMSE Wavelet Shrinkage. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 606–616. [CrossRef]
23. Argenti, F.; Lapini, A.; Bianchi, T.; Alparone, L. A Tutorial on Speckle Reduction in Synthetic Aperture Radar Images. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–35. [CrossRef]
24. Li, J.; Yu, Z.; Yu, L.; Cheng, P.; Chen, J.; Chi, C. A Comprehensive Survey on SAR ATR in Deep-Learning Era. *Remote Sens.* **2023**, *15*, 1454. [CrossRef]
25. Zhou, Z.; Cui, Z.; Tang, K.; Tian, Y.; Pi, Y.; Cao, Z. Gaussian Meta-Feature Balanced Aggregation for Few-Shot Synthetic Aperture Radar Target Detection. *ISPRS J. Photogramm. Remote Sens.* **2024**, *208*, 89–106. [CrossRef]
26. Ying, L.; Miao, D.; Zhang, Z. A Robust One-Stage Detector for SAR Ship Detection with Sequential Three-Way Decisions and Multi-Granularity. *Inf. Sci.* **2024**, *667*, 120436. [CrossRef]
27. Shen, Y.F.; Gao, Q. DS-YOLO: A SAR Ship Detection Model for Dense Small Targets. *Radioengineering* **2025**, *34*, 407–421. [CrossRef]
28. Xu, X.; Zhang, X.; Zhang, T. Lite-YOLOv5: A Lightweight Deep Learning Detector for On-Board Ship Detection in Large-Scene Sentinel-1 SAR Images. *Remote Sens.* **2022**, *14*, 1018. [CrossRef]

29. Yasir, M.; Liu, S.; Pirasteh, S.; Xu, M.; Sheng, H.; Wan, J.; de Figueiredo, F.A.P.; Aguilar, F.J.; Li, J. YOLOShipTracker: Tracking Ships in SAR Images Using Lightweight YOLOv8. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *134*, 104137. [CrossRef]
30. Dong, J.; Feng, J.; Tang, X. OptiSAR-Net: A Cross-Domain Ship Detection Method for Multisource Remote Sensing Data. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4709311. [CrossRef]
31. Xu, W.; Guo, Z.; Huang, P.; Tan, W.; Gao, Z. Towards Efficient SAR Ship Detection: Multi-Level Feature Fusion and Lightweight Network Design. *Remote Sens.* **2025**, *17*, 2588. [CrossRef]
32. He, S. Multiscale Task-Decoupled Oriented SAR Ship Detection Based on a Size-Aware Balanced Strategy. *Remote Sens.* **2025**, *17*, 2257. [CrossRef]
33. Wu, B.; Liu, C.; Chen, Z.; Zhang, S.; Chen, J. A Novel Star Feature Decoupling Network for Multiscale SAR Image Ship Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2026**. [CrossRef]
34. Sun, Z.; Leng, X.; Zhang, X.; Zhou, Z.; Xiong, B.; Ji, K.; Kuang, G. Arbitrary-Direction SAR Ship Detection Method for Multiscale Imbalance. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5208921. [CrossRef]
35. Chen, B.; Xue, F.; Song, H. Lightweight Transformer Detector for SAR Ship Detection. *IEEE Trans. Pattern Anal. Mach. Intelling* **2024**, *16*, 237. [CrossRef]
36. Zhang, Y.; Chi, J.; Yang, G.; Chen, C.; Yu, T. SAR-NanoShipNet: A Scale-Adaptive Network for Robust Small Ship Detection in SAR Imagery. *ISPRS J. Photogramm. Remote Sens.* **2026**, *232*, 262–279. [CrossRef]
37. Zhao, S.; Luo, Y.; Zhang, T.; Guo, W.; Zhang, Z. A Domain Specific Knowledge Extraction Transformer Method for Multisource Satellite-Borne SAR Images Ship Detection. *ISPRS J. Photogramm. Remote Sens.* **2023**, *198*, 16–29. [CrossRef]
38. Yang, Y.; Chen, J.; Sun, L.; Zhou, Z.; Huang, Z.; Wu, B. Unsupervised Domain-Adaptive SAR Ship Detection Based on Cross-Domain Feature Interaction and Data Contribution Balance. *Remote Sens.* **2024**, *16*, 420. [CrossRef]
39. Li, D.; Liang, Q.; Liu, H.; Liu, Q.; Liu, H.; Liao, G. A Novel Multidimensional Domain Deep Learning Network for SAR Ship Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5203213. [CrossRef]
40. Ding, X.; Zhang, X.; Han, J.; Ding, G. Scaling up Your Kernels to 31 × 31: Revisiting Large Kernel Design in CNNs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022*; IEEE: New York, NY, USA, 2022; pp. 11963–11975. [CrossRef]
41. Wang, W.; Dai, J.; Chen, Z.; Huang, Z.; Li, Z.; Zhu, X.; Hu, X.; Lu, T.; Lu, L.; Li, H.; et al. InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022*; IEEE: New York, NY, USA, 2022; pp. 14408–14419.
42. Zhang, X.; Liu, C.; Yang, D.; Song, T.; Ye, Y.; Li, K.; Song, Y. RFAConv: Innovating Spatial Attention and Standard Convolutional Operation. *arXiv* **2024**, arXiv:2304.03198. [CrossRef]
43. Cui, Y.; Tao, Y.; Bing, Z.; Ren, W.; Gao, X.; Cao, X.; Huang, K.; Knoll, A. Selective Frequency Network for Image Restoration. In Proceedings of the Eleventh International Conference on Learning Representations (ICLR), Kigali, Rwanda, 1–5 May 2023.
44. Pham, C.; Nguyen, V.; Le, T.; Phung, D.; Carneiro, G.; Do, T. Frequency Attention for Knowledge Distillation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2024*; IEEE: New York, NY, USA, 2024; pp. 2277–2286. [CrossRef]
45. Chen, J.; Duanmu, C.; Long, H. Large Kernel Frequency-Enhanced Network for Efficient Single Image Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 17–18 June 2024*; IEEE: New York, NY, USA, 2024; pp. 6317–6326. [CrossRef]
46. Ni, K.; Wang, P.; Zheng, Z.; Zhong, Y. Complex-valued mix transformer for SAR ship detection. *ISPRS J. Photogramm. Remote Sens.* **2026**, *231*, 1–16. [CrossRef]
47. Tian, Z.; Wu, Q.; Xu, J.; Zhong, Y. FCOS: Fully Convolutional One-Stage Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019*; IEEE: New York, NY, USA, 2019; pp. 9627–9636. [CrossRef]
48. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. In Proceedings of the 34 International Conference on Neural Information Processing Systems (NeurIPS), Vancouver, BC, Canada, 6–12 December 2020.
49. Yang, Z.; Liu, S.; Hu, H.; Wang, L.; Lin, S. RepPoints: Point Set Representation for Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October –2 November 2019*; IEEE: New York, NY, USA, 2019; pp. 9657–9666. [CrossRef]
50. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as Points. *arXiv* **2019**, arXiv:1904.07850. [CrossRef]
51. Wang, W.; Chen, K.; Kim, P.; Yoon, K.; Liu, Z.; Joo, K.; Liu, Y.-H. Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction Without Convolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021*; IEEE: New York, NY, USA, 2021; pp. 568–578. [CrossRef]
52. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017*; IEEE: New York, NY, USA, 2017; pp. 2980–2988.

53. Feng, C.; Zhong, Y.; Gao, Y.; Scott, M.R.; Huang, W. TOOD: Task-Aligned One-Stage Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021*; IEEE: New York, NY, USA, 2021; pp. 3490–3499. [CrossRef]
54. Chen, Z.; Yang, C.; Li, Q.; Zhao, F.; Zha, Z.; Wu, F. Dicentang: Your Dense Object Detector. In *Proceedings of the 29th ACM International Conference on Multimedia (ACM MM), Virtual Event, China, 20–24 October 2021*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 4939–4948.
55. Chen, Q.; Wang, Y.; Yang, T.; Zhang, X.; Cheng, J.; Sun, J. You Only Look One-Level Feature. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021*; IEEE: New York, NY, USA, 2021; pp. 13039–13048. [CrossRef]
56. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430. [CrossRef]
57. Zhang, M.; Zhu, Y.; Li, L.; Guo, J.; Liu, Z.; Li, Y. S4Det: Breadth and Accurate Sine Single-Stage Ship Detection for Remote Sense SAR Imagery. *Remote Sens.* **2025**, *17*, 900. [CrossRef]
58. Liu, Z.; Mao, H.; Wu, C.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022*; IEEE: New York, NY, USA, 2022; pp. 11976–11986.
59. Woo, S.; Debnath, S.; Hu, R.; Chen, X.; Liu, Z.; Kweon, I.S.; Xie, S. ConvNeXt V2: Co-Designing and Scaling ConvNets with Masked Autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023*; IEEE: New York, NY, USA, 2023; pp. 16133–16142.
60. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 213–229. [CrossRef]
61. Zhang, M.; Li, Y.; Guo, J.; Li, Y.; Gao, X. BurgsVO: Burgs-Associated Vertex Offset Encoding Scheme for Detecting Rotated Ships in SAR Images. *Remote Sens.* **2025**, *17*, 388. [CrossRef]
62. Liu, S.; Li, F.; Zhang, H.; Yang, X.; Qi, X.; Su, H.; Zhu, J.; Zhang, L. DAB-DETR: Dynamic Anchor Boxes are Better Queries for DETR. In Proceedings of the International Conference on Learning Representations, Online, 25–29 April 2022.
63. Meng, D.; Chen, X.; Fan, Z.; Zeng, G.; Li, H.; Yuan, Y.; Sun, L.; Wang, J. Conditional DETR for Fast Training Convergence. In *Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021*; IEEE: New York, NY, USA, 2021; pp. 3651–3660. [CrossRef]
64. Zhou, J.; Xiao, C.; Peng, B.; Liu, Z.; Liu, L.; Liu, Y.; Li, X. DiffDet4SAR: Diffusion-Based Aircraft Target Detection Network for SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 4007905. [CrossRef]
65. Wang, T.; Zeng, Z.; Zhou, S.; Xu, Q. A Multi-Scale Discrete Feature Enhancement Network with Augmented Reversible Transformation for SAR Automatic Target Recognition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 5135–5156. [CrossRef]
66. Dai, Y.; Zou, M.; Li, Y.; Li, X.; Ni, K.; Yang, J. DenoDet: Attention as Deformable Multi-Subspace Feature Denoising for Target Detection in SAR Images. *arXiv* **2024**, arXiv:2406.02833. [CrossRef]
67. Zhang, M.; Yang, Z.; Guo, J.; Li, Y. PDE-Guided Diverse Feature Learning for SAR Rotated Ship Detection. *Remote Sens.* **2025**, *17*, 2998. [CrossRef]
68. Zhang, S.; Chi, C.; Yao, Y.; Lei, Z.; Li, S.Z. Bridging the Gap Between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020*; IEEE: New York, NY, USA, 2020; pp. 9759–9768. [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.