

Article

Modulation and Perturbation in Frequency Domain for SAR Ship Detection

Mengqin Fu ¹, Wencong Zhang ^{2,*}, Xiaochen Quan ², Dahu Shi ², Luowei Tan ¹, Jia Zhang ¹, Yinghui Xing ¹ and Shizhou Zhang ¹

¹ National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China; mengqinfu@mail.nwpu.edu.cn (M.F.); tanluowei@mail.nwpu.edu.cn (L.T.); zzzhja@mail.nwpu.edu.cn (J.Z.); xyh_7491@nwpu.edu.cn (Y.X.); szzhang@nwpu.edu.cn (S.Z.)

² Hikrobot Co., Ltd., Hangzhou 310051, China; quanxiaochen@hikrobotics.com (X.Q.); shidahu@hikrobotics.com (D.S.)

* Correspondence: zhangwencong@hikrobotics.com

Highlights

What are the main findings?

- Propose a Frequency-aware Modulation and Perturbation Network (FMP-Net) integrating adaptive wavelet attention and Frequency-Guided Perturbation, which selectively enhances ship high-frequency features while suppressing speckle noise and background interference.
- Achieve state-of-the-art performance on four benchmark datasets (LS-SSDD-v1.0, SSDD, SAR-Ship-Dataset, AIR-SARShip2.0) with a lightweight design (2.78M parameters), outperforming 18 existing methods, especially in complex nearshore and sea clutter scenes.

What are the implications of the main findings?

- Provides a lightweight, high-performance practical solution for SAR ship detection in complex sea areas like nearshore and sea clutter scenarios, which can improve the efficiency of practical tasks including maritime surveillance and maritime security.
- Verifies the effectiveness of the technical idea combining “frequency-domain modulation + perturbation” with attention mechanism in SAR image denoising and target feature enhancement, offering new technical references for subsequent algorithm design of SAR target detection.
- The proposed FMP-Net can be easily integrated into existing object detection frameworks with minimal computational cost, providing a versatile tool for enhancing SAR ship detection across various applications.



Academic Editor: Zhenwei Shi

Received: 28 November 2025

Revised: 2 January 2026

Accepted: 8 January 2026

Published: 20 January 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Abstract

Synthetic Aperture Radar (SAR) has unique advantages in ship monitoring at sea due to its all-weather imaging capability. However, its unique imaging mechanism presents two major challenges. First, speckle noise in the frequency domain reduces the contrast between the target and the background. Second, side-lobe scattering blurs the ship outline, especially in nearshore complex scenes, and strong scattering characteristics make it difficult to separate the target from the background. The above two challenges significantly limit the performance of tailored CNN-based detection models in optical images when applied directly to SAR images. To address these challenges, this paper proposes a modulation and perturbation mechanism in the frequency domain based on a lightweight CNN detector.

Specifically, the wavelet transform is firstly used to extract high-frequency features in different directions, and feature expression is dynamically adjusted according to the global statistical information to realize the selective enhancement of the ship edge and detail information. In terms of frequency-domain perturbation, a perturbation mechanism guided by frequency-domain weight is introduced to effectively suppress background interference while maintaining key target characteristics, which improves the robustness of the model in complex scenes. Extensive experiments on four widely adopted benchmark datasets, namely LS-SSDD-v1.0, SSDD, SAR-Ship-Dataset, and AIR-SARShip-2.0, demonstrate that our FMP-Net significantly outperforms 18 existing state-of-the-art methods, especially in complex nearshore scenes and sea surface interference scenes.

Keywords: SAR ship target detection; frequency-domain adaptation; modulation; feature perturbation

1. Introduction

Synthetic Aperture Radar (SAR) is crucial for maritime ship detection, ocean observation, and maritime security because of its capability to provide images regardless of weather or time of day [1]. SAR is particularly beneficial for detecting maritime entities as it can penetrate challenging weather conditions like clouds, rain, and fog, allowing for continuous marine surveillance [2]. Nonetheless, SAR's specialized imaging method presents various hurdles, complicating the ship detection process [3]. Firstly, speckle noise in SAR imagery poses a significant challenge. This noise originates from the coherent interference between radar and scattered waves on target surfaces [4], spreading widely across the frequency domain. The presence of speckle noise introduces random grainy textures in images, diminishing the contrast between targets and their backgrounds, thereby complicating detection [5]. As depicted in Figure 1a, background noise from waves, ocean currents, and sea surface clutter makes small targets more susceptible to noise disruption or being obfuscated by clutter. Secondly, scattering leakage from side lobes can render the ship's external outlines incomplete and indistinct. As illustrated in Figure 1b,c, differentiation between target and background becomes more challenging in complex inshore environments due to intense side-lobe leakage and robust scattering from coastal and port infrastructures.

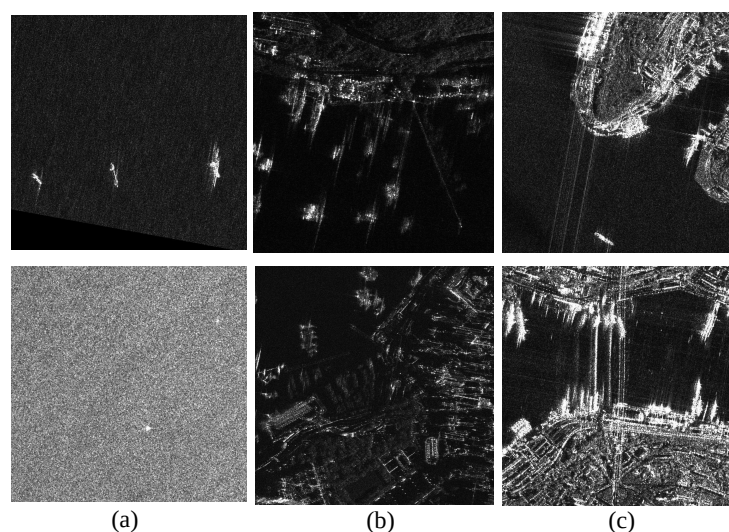


Figure 1. Several typical SAR target detection challenges: (a) sea clutter interference, (b) strong side-lobe interference leading to shape and edge blur, (c) complex scenes near the shore.

Traditional ship detection methods in SAR imagery, such as the constant false alarm rate (CFAR) statistical approaches [6–8], frequently rely on predetermined thresholds based on a specific statistical model for the background noise. These methods function well in largely uniform backgrounds, but in more intricate maritime and coastal regions, noise and target distribution uncertainties make target–background differentiation problematic. Object detection models leveraging Convolutional Neural Networks (CNNs), such as the Faster R-CNN [9] and YOLO versions [10–16], can independently extract image features through comprehensive training and have been proven effective in object detection tasks. However, as illustrated in Figure 2, there is a significant discrepancy between the imaging qualities of SAR and optical images [3]. This disparity often results in difficulties when these models are applied, as they encounter issues with speckle noise and complex background interference, leading to reduced performance. These deep learning models, in particular, find it challenging to distinguish ship targets from other highly scattered objects in complex settings and have trouble adapting to variable backgrounds, especially near ports and coastlines, resulting in increased false positive rates.

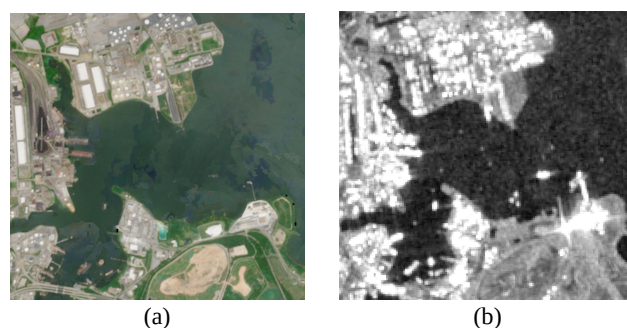


Figure 2. Examples of optical and SAR images from the same view. (a) Visible image. (b) SAR image.

Incorporating frequency-domain information offers a powerful means of addressing these challenges. Within SAR imagery, ship targets' edge and detailed features typically manifest as high-frequency components, whereas the background is primarily located in the lower-frequency range. Consequently, frequency-domain analysis can adeptly separate the target's frequency elements from the background, thereby enhancing target detection accuracy. Nonetheless, speckle noise infiltrates not only the low-frequency regions but also disrupts the image details within the high-frequency zones. This noise reduces the contrast between targets and backgrounds, compounding the difficulty of object detection. The presence of high-frequency noise necessitates not just attention to the targets' edges and details during high-frequency feature extraction but also requires effective discrimination between beneficial high-frequency data and noise. Although conventional frequency domain filtering techniques, such as high-pass filtering, succeed in isolating high-frequency components, they fail to differentiate between target characteristics and noise, potentially escalating high-frequency noise and impairing the efficacy of the detector.

This paper introduces a Frequency-aware Modulation and Perturbation Network (FMP-Net) aimed at improving SAR ship detection. As shown in Figure 3, the core of FMP-Net is the two successive Frequency-aware Modulation and Perturbation (FMP) stages, composed of two crucial components: the multiscale Frequency-aware Modulation and Perturbation (MS-FMP) module and the Adaptive Wavelet Frequency-aware Downsampling (AWFDown) module, whose detailed architectures are illustrated in Figure 4. Within the MS-FMP module, the input feature map initially undergoes a multiscale convolution process. Subsequently, these features flow into the Adaptive Frequency-aware Modulation and Perturbation (AFMP) module for further refinement. The AFMP module features two branches: one for capturing and adjusting high-frequency features, and the other

for diminishing background interference. The first branch employs an Adaptive Wavelet Frequency-aware Attention (AWFA) module to detect high-frequency data across various directions. Concurrently, to prevent undue enhancement of high-frequency noise, a variance modulation strategy is employed, adjusting high-frequency features based on their variance to minimize false detections in complex backgrounds. The opposing branch houses the Frequency-Guided Perturbation (FGP) module, which perturbs feature maps within the same batch. Unlike random perturbation, this method uses frequency domain weights from the AWFA module to guide perturbation, maintaining essential details and textures while suppressing less critical features and improving the model noise resistance, as illustrated in Figure 5. To the best of our knowledge, this is the first time feature perturbation has been introduced into the field of SAR target detection. In the AWFDn module, the AWFA module is additionally integrated to preserve high-frequency edges and textures during downsampling, ensuring critical features remain intact while reducing resolution. The contributions of the paper can be summarized in three key aspects:

- We introduce an innovative Frequency-aware Modulation and Perturbation Network that effectively leverages the frequency domain characteristics of SAR images to thoroughly explore target features, thereby enhancing the precision and resilience of ship detection.
- Our developed Frequency-aware Modulation and Perturbation module employs a dual-branch structure to improve the model's resistance to noise: the first branch is designed to modulate crucial frequency components through a tailored AWFA module, while the second branch perturbs the feature maps guided by frequency weights generated by AWFA.
- Our approach significantly surpasses baseline methods and attains state-of-the-art results across four SAR ship target detection benchmark datasets, namely LS-SSDD-v1.0, SSDD, SAR-Ship-Dataset, and AIR-SARShip-2.0. Moreover, the method can be easily integrated into existing object detection frameworks with minimal computational cost.

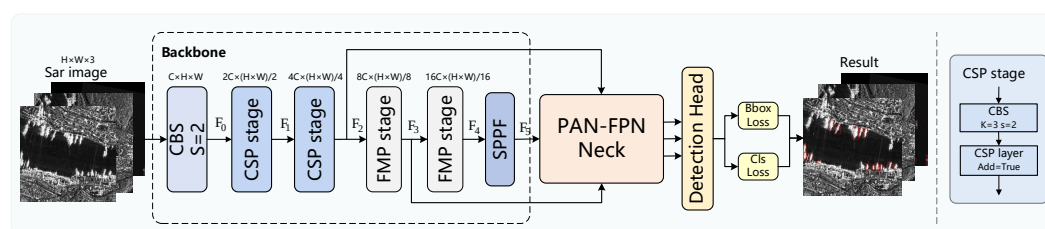


Figure 3. The overall framework of our proposed FMP-Net.

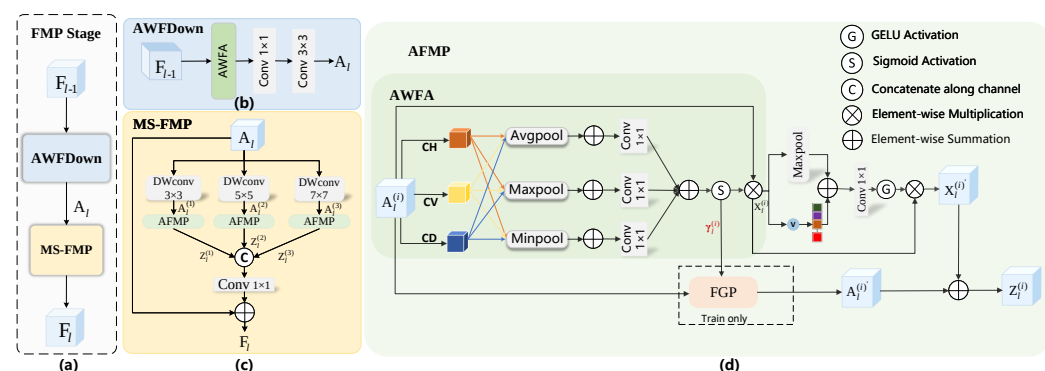


Figure 4. Overall framework of our proposed Frequency-aware Modulation and Perturbation (FMP) stage. (a) The main architecture. (b) Adaptive Wavelet Frequency-aware Downsampling (AWFDn) module. (c) Mutiscale Frequency-aware Modulation and Perturbation (MS-FMP) module. (d) Adaptive Frequency-aware Modulation and Perturbation (AFMP) module.

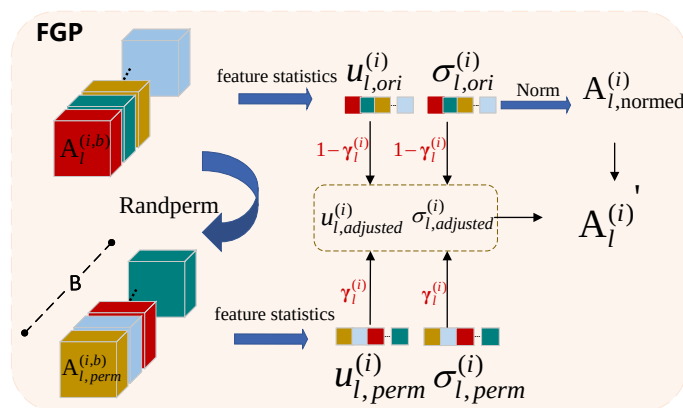


Figure 5. The detailed architecture of our proposed Frequency-Guided Perturbation (FGP) module.

The remainder of this paper is organized as follows: Section 2 provides a summary of the related work discussed in this study. Section 3 details the components of the proposed approach. As discussed in Section 4, we performed comprehensive experiments on four widely adopted benchmark datasets to demonstrate the advantages of our method and evaluate the contributions of its individual components. Finally, Section 5 presents the conclusions of this study.

2. Related Work

2.1. SAR Ship Object Detection

Due to significant differences in feature distribution and noise characteristics between SAR images and optical images, the effect of the general target detection model applied directly to the SAR images is sub-optimal. To this end, researchers have improved these models to adapt to the characteristics of SAR images. For example, Fu et al. [17] proposed an anchor-free detection framework for multiscale SAR ship detection to eliminate the influence of anchors by directly learning the encoded bounding boxes. Yang et al. [18] introduced a coordinate attention module (CoAM) into the backbone for multiscale ship detection with complex backgrounds in massive SAR images. Hong et al. [19] improved YOLOv3 [10] according to the characteristics of the ships and introduced a Gaussian model to output the uncertainty of each predicted bounding box to improve detection accuracy. Hu et al. [20] proposed the Feature Interaction Network (FINet), which considers the interaction of different subtasks at different scales from the object level and the pixel level to improve the localization and classification ability in complex scenes. Zha et al. [21] proposed a network based on Salient Region Extraction (SRE) and Multi-branch Attention (MBA) to enhance the saliency of ships and improved the detection accuracy of small ships. Muhammad et al. [22] combined a CNN and non-linear regression to extract ship geometric features from SAR images and improved Mask R-CNN [23] and ViTDet [24] models to generate high-quality object masks for accurate localization. Shen et al. [25] used an anchor-free detector, decomposed large kernel convolutions to capture global information and long-range dependencies, adaptively selected convolution kernels based on target features, and improved multiscale feature representation. Muhammad et al. [26] proposed SwinYOLOv7 by combining YOLOv7 [27], a Feature Pyramid Network (FPN) [28], and the Swin Transformer to reduce the influence of background clutter. Besides intensity-based SAR detectors, polarimetric and decomposition-based methods have also been explored to exploit scattering information for ship-related analysis. Zhang et al. [29] proposed a PolSAR ship detection metric based on sub-look decomposition and $H/A/\alpha$ analysis to suppress sea clutter. Wang et al. [30] developed a PSO-based fine polarimetric decomposition with explicit physical meaning for ship scattering characterization. These

physics-driven approaches typically rely on full-polarimetric measurements and explicit decomposition procedures. In contrast, our method targets conventional SAR intensity images and improves ship detection robustness within a lightweight CNN detector under speckle noise and complex backgrounds.

In addition, some recent works [31–33] have investigated the use of frequency domain information to enhance the feature enhancement of the SAR image target. In contrast, we propose a Frequency-aware Modulation and Perturbation strategy that not only selectively enhances high-frequency target features through adaptive modulation but also introduces Frequency-Guided Perturbation to suppress background interference, enabling more robust detection in complex SAR scenes.

2.2. SAR Target Detection Based on Frequency Domain

Considering the frequency characteristics of SAR images, some studies have begun to focus on the application of frequency domain features for SAR target detection. Li et al. [34] manually constructed the rotation-invariant frequency domain feature of the ship target with the polar Fourier transform to improve the detection of multiscale ships in complex SAR scenes. Qiao et al. [35] proposed Multi-frequency Aggregation (MFA), which uses Discrete Cosine Transform (DCT) to aggregate frequency information and suppresses background interference by selecting the optimal frequency response feature. Zhang et al. [31] introduced the frequency attention mechanism (FAM) to suppress the interference of sea clutter through the end-to-end adaptive processing of frequency-domain information from SAR images. Wang et al. [32] used the Haar wavelet transform to convert the original image into the frequency domain and obtained the spatial and frequency characteristics of the target ship to effectively separate the key features and useless noise of the SAR image. However, the above frequency-domain methods do not fully use high-frequency information. For SAR intensity imagery, high-frequency components simultaneously encode discriminative ship edges and textures as well as speckle- and background-induced interference, so indiscriminate high-frequency enhancement tends to amplify noise and reduce reliability. Therefore, simply amplifying the high-frequency information will cause the background noise to be amplified at the same time, thus reducing the detection accuracy and reliability. The variance modulation mechanism introduced in this paper adaptively adjusts the high-frequency features through global statistical information to achieve the goal of selectively enhancing high-frequency features and suppressing background noise. In summary, our FMP module differs from prior frequency-domain-based detectors in three aspects. First, we model directional high-frequency responses via wavelet-aware weighting, rather than using a single global frequency representation. Second, we introduce variance modulation to adaptively balance enhancement and suppression in the high-frequency branch, which avoids indiscriminate amplification of noisy responses. Third, beyond deterministic enhancement, we incorporate frequency-guided feature perturbation to improve robustness against speckle noise and complex background interference.

2.3. Feature Perturbation

In the realm of object detection, feature perturbation has been proven to effectively boost a model's ability to generalize [36,37]. Modifying input features prevents the model from heavily depending on specific local traits, thereby enhancing its resilience across varying conditions [36–44]. For instance, Zhou et al. [42] introduced AMSP Vortex Convolution (AMSP-VCONV) to alter the noise pattern and strengthen feature extraction. Fan et al. [43] devised the Normalized Perturbation (NP) method to create diverse latent styles by modifying channel statistics of low-level source-domain features, enabling the model to recognize various latent domains. Zhou et al. [44] introduced Mixstyle, which in training combines

feature statistics from two random samples in one pass to improve domain generalization performance. While feature perturbation based on spatial features has been effective in detecting targets in optical images, its usage in SAR image target detection remains limited. SAR images, often complicated by speckle noise and intricate backgrounds, face the challenge of noise interference, which may result in false detections and missed targets. This study presents a feature perturbation module founded on high-frequency information enhancement, which further augments the model's adaptability in cluttered and noisy settings.

3. Approach

3.1. Overall Framework

The overall architecture of our proposed FMP-Net is depicted in Figure 3. It comprises three primary components: backbone, neck, and detection head. In this work, we adopted YOLOv8n as the baseline detection framework. Initially, SAR images, denoted as $\mathbf{I} \in \mathbb{R}^{H \times W}$, are introduced as input data and processed by the backbone network. This backbone is made up of two CSP stages and two Frequency-aware Modulation and Perturbation (FMP) stages. The CSP stages handle the extraction of low-level features, whereas the FMP stages are responsible for capturing higher-level semantic features. Each FMP stage consists of an Adaptive Wavelet Frequency-aware Downsampling (AWFDown) module and a Multiscale Frequency-aware Modulation and Perturbation (MS-FMP) module, as illustrated in Figure 4. The AWFDown module maintains the image's high-frequency information during downsampling via wavelet decomposition, ensuring that essential details are retained despite the reduction in the map's size. The MS-FMP module generates multiscale features through the Adaptive Frequency-aware Modulation and Perturbation (AFMP) module, primarily comprising an Adaptive Wavelet Frequency-aware Attention (AWFA) module and a variance modulation module. The AWFA module captures edge and detail information of the image to ensure precise contour extraction of targets, whereas the variance modulation utilizes global statistical characteristics to dynamically adjust high-frequency features, preventing false detections in complex environments or with significant noise. Moreover, the Frequency-Guided Perturbation (FGP) module within the AFMP module aids in mitigating background interference on the target by introducing dynamic disturbances during training, thereby enhancing detection robustness. Subsequently, the SPPF [45] module further refines these features to bolster computational efficiency and eliminate redundant data. The PAN-FPN [46] structure is employed to merge multilevel features from the backbone, enabling the model to detect both detailed and global features effectively. Ultimately, the detection head utilizes features obtained by the neck to perform object classification and localization, providing the final detection results.

3.2. Frequency-Aware Modulation and Perturbation (FMP)

As shown in Figure 4a, the FMP stage consists of two main parts: the MS-FMP and AWFDown modules, which successively process the input features in order to retain high-frequency information while downsampling, selectively enhancing and perturbing high-frequency features. Given the input feature map $\mathbf{F}_{l-1} \in \mathbb{R}^{\frac{H}{2^{l-1}} \times \frac{W}{2^{l-1}} \times C_{l-1}}$, $l \in [3, 4]$ denotes the number of stages that are first downsampled by the AWFDown module to obtain the feature map $\mathbf{A}_l \in \mathbb{R}^{\frac{H}{2^l} \times \frac{W}{2^l} \times C_l}$. Then, $\mathbf{A}_l$ is processed by the MS-FMP module, which further enhances and perturbs high-frequency features, resulting in the output feature map $\mathbf{F}_l \in \mathbb{R}^{\frac{H}{2^l} \times \frac{W}{2^l} \times C_l}$, i.e.,

$$\mathbf{F}_l = \text{MS_FMP}(\text{AWFDown}(\mathbf{F}_{l-1})). \quad (1)$$

Adaptive Wavelet Frequency-Aware Downsampling (AWFDown). Due to the possible loss of important information in detail during the sampling operation, we introduce

AWFDown to effectively preserve high-frequency information during the downsampling process. As shown in Figure 4b, AWFDown consists of AWFA, $\text{Conv}_{3 \times 3}(\cdot)$, and $\text{Conv}_{1 \times 1}(\cdot)$:

$$\mathbf{A}_l = \text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(\text{AWFA}(\mathbf{F}_{l-1}))), \quad (2)$$

where the AWFA module can enhance the edges and details in the feature map, so as to retain important detail information in the subsequent downsampling process. $\text{Conv}_{1 \times 1}(\cdot)$ represents a convolutional layer with a kernel size of 1×1 , and $\text{Conv}_{3 \times 3}(\cdot)$ represents a convolutional layer with a kernel size of 3×3 . In Section 3.3, we discuss the implementation of the AWFA module in detail.

Mutiscale Frequency-Aware Modulation and Perturbation (MS-FMP). As mentioned in Section 1, unlike general target detection, SAR target detection is susceptible to strong scattering and speckle noise, making it difficult to distinguish the target from the background. In order to solve the problem of the ambiguous boundary between the background and the object, as shown in Figure 4c, we introduce the MS-FMP module to capture multiscale texture features. Formally, the MS-FMP module within the l -th stage can be represented mathematically as

$$\begin{aligned} \mathbf{Z}_l^{(i)} &= \text{AFMP}(\text{DWConv}_{k_i \times k_i}(\mathbf{A}_l)), \\ k_i &= \{3, 5, 7\}, i = \{1, 2, 3\}, \\ \mathbf{F}_l &= \text{Conv}_{1 \times 1}(\text{Concat}(\mathbf{Z}_l^{(1)}, \mathbf{Z}_l^{(2)}, \mathbf{Z}_l^{(3)})) + \mathbf{A}_l. \end{aligned} \quad (3)$$

3.3. Adaptive Frequency-Aware Modulation and Perturbation (AFMP) Module

In SAR images, important information about the target, such as edge and texture, often appears in the form of high-frequency features, and background noise may also appear as high-frequency interference. Therefore, as shown in Figure 4d, we designed the AFMP module to achieve selective enhancement of high-frequency features and suppression of background noise.

Adaptive Wavelet Frequency-Aware Attention Module (AWFA). As shown in Figure 4d, the AWFA module first applies a one-level Haar wavelet transform to the input feature map $\mathbf{A}_l^{(i)} \in \mathbb{R}^{\frac{H}{2^l} \times \frac{W}{2^l} \times C_l}$, where $i \in \{1, 2, 3\}$ denotes the i -th scale, to extract high-frequency information that is indicative of boundaries and structural details. Through Haar wavelet decomposition, we retain only the three directional high-frequency components: the horizontal detail component $\mathbf{X}_l^{LH} \in \mathbb{R}^{\frac{H}{2^{l+1}} \times \frac{W}{2^{l+1}} \times C_l}$, the vertical detail component $\mathbf{X}_l^{HL} \in \mathbb{R}^{\frac{H}{2^{l+1}} \times \frac{W}{2^{l+1}} \times C_l}$, and the diagonal detail component $\mathbf{X}_l^{HH} \in \mathbb{R}^{\frac{H}{2^{l+1}} \times \frac{W}{2^{l+1}} \times C_l}$. The rationale for using $\mathbf{X}_l^{LH}$, $\mathbf{X}_l^{HL}$, and $\mathbf{X}_l^{HH}$ jointly is that ship boundaries, wakes, and corner-type scattering structures may appear at arbitrary orientations. Accordingly, different boundary segments can yield stronger responses in different directional components, and using only one component may miss edges with other orientations. By aggregating these three components, AWFA forms a more complete edge-enhancement representation, which facilitates target-background separation in both offshore scenes and nearshore scenarios with complex clutter. This behavior is illustrated in Figure 6, where the combined magnitude of LH+HL+HH highlights ship boundaries more clearly than any single directional component.

These high-frequency components can be obtained through convolution with the following Haar filter kernels:

$$f_{LH} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, f_{HL} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, f_{HH} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad (4)$$

where f_{LH} , f_{HL} , and f_{HH} are used to extract high-frequency information in horizontal, vertical, and diagonal directions, respectively. In implementation, the Haar transform is realized by fixed 2×2 Haar filter kernels with a stride of 2 applied over the entire feature map, and thus, the transform window is inherently 2×2 for a one-level decomposition. This design does not introduce an additional patch-wise window-size hyper-parameter; instead, it performs dense local filtering across the whole feature map with minimal computational overhead. By applying these filters with a stride of 2, high-frequency components in the horizontal, vertical, and diagonal directions can be obtained from the input feature map $\mathbf{A}_l^{(i)}$,

$$[\mathbf{X}_l^{LH}, \mathbf{X}_l^{HL}, \mathbf{X}_l^{HH}] = \text{Conv}([f_{LH}, f_{HL}, f_{HH}], \mathbf{A}_l^{(i)}). \quad (5)$$

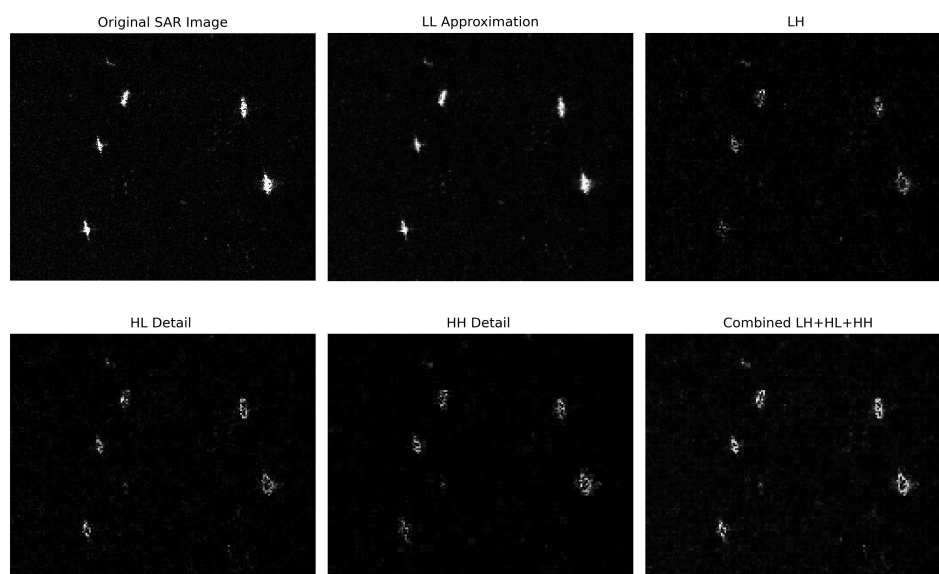


Figure 6. Visualization of one-level 2D Haar DWT subbands on a representative SAR image. LL denotes the low-frequency approximation. LH, HL, and HH denote the directional high-frequency detail subbands.

To capture frequency features at different scales, we apply multiscale pooling to the high-frequency components $\mathbf{X}_l^{LH}$, $\mathbf{X}_l^{HL}$, and $\mathbf{X}_l^{HH}$. Specifically,

$$\begin{aligned} \mathbf{X}_{l,avg}^{(i)} &= \sum_{h \in \{LH, HL, HH\}} \text{AvgPool}(\mathbf{X}_l^h), \\ \mathbf{X}_{l,max}^{(i)} &= \sum_{h \in \{LH, HL, HH\}} \text{MaxPool}(\mathbf{X}_l^h), \\ \mathbf{X}_{l,min}^{(i)} &= \sum_{h \in \{LH, HL, HH\}} \text{MinPool}(\mathbf{X}_l^h), \end{aligned} \quad (6)$$

where $\mathbf{X}_{l,avg}^{(i)}$, $\mathbf{X}_{l,max}^{(i)}$, and $\mathbf{X}_{l,min}^{(i)} \in \mathbb{R}^{1 \times 1 \times C_l}$ represent the average, maximum, and minimum pooling results, respectively. These operations smooth, emphasize, and suppress high-frequency features to enhance texture details, highlight structural edges, and reduce background noise. The pooled features are then fused by applying a 1×1 convolution to each pooled output and summing the results, as follows:

$$\begin{aligned} \lambda_l^{(i)} &= \sigma \left(\sum_{p \in \{avg, max, min\}} \text{Conv}_{1 \times 1}(\mathbf{X}_{l,p}^{(i)}) \right), \\ \mathbf{X}_l^{(i)} &= \lambda_l^{(i)} \otimes \mathbf{A}_l^{(i)}, \end{aligned} \quad (7)$$

where $\mathbf{X}_l^{(i)} \in \mathbb{R}^{\frac{H}{2^l} \times \frac{W}{2^l} \times C_l}$ denotes the output of the AWFA module. The weight of the frequency domain $\lambda_l^{(i)} \in \mathbb{R}^{1 \times 1 \times C_l}$, obtained through the sigmoid activation function σ , dynamically modulates the strength of the high-frequency features in different regions of the feature map. “ $\otimes$ ” represents element-wise broadcasted multiplication. This adaptive weighting mechanism enables the model to selectively enhance the high-frequency details in the target regions.

Variance modulation. In SAR images, high-frequency information not only includes target edges and details but also contains a significant amount of noise. Direct enhancement of all high-frequency features may lead to excessive amplification of noise. Variance modulation dynamically adjusts the expression of high-frequency information by combining the global statistical information of the feature map with local salient features.

Firstly, by calculating the variance of the feature map, we can quantify the global variation range of each channel of the feature map, which can reflect the strength of the high-frequency features in the image.

$$\begin{aligned} \mathbf{v}_l^{(i)} &= \text{Var}(\mathbf{X}_l^{(i)}) \\ &= \frac{1}{HW} \sum_{p=1}^H \sum_{q=1}^W \left(\mathbf{X}_{l,pq}^{(i)} - \mu(\mathbf{X}_l^{(i)}) \right)^2, \\ \mu(\mathbf{X}_l^{(i)}) &= \frac{1}{HW} \sum_{p=1}^H \sum_{q=1}^W \mathbf{X}_{l,pq}^{(i)}. \end{aligned} \quad (8)$$

where $\mu(\mathbf{X}_l^{(i)})$ represents the mean of the feature map $\mathbf{X}_l^{(i)}$. Next, to capture prominent local high-frequency features, we apply adaptive maxpooling to extract significant local details. We then combine the global variance $\mathbf{v}_l^{(i)} \in \mathbb{R}^{1 \times 1 \times C_l}$ with the locally pooled feature through a linear combination, weighted by learnable parameters α and β , to balance the influence of global and local high-frequency information. The combined features undergo a linear transformation with a learnable weight matrix $\mathcal{W}_l^{(i)} \in \mathbb{R}^{1 \times 1 \times C_l}$ and are then passed through a GELU activation function. The modulation process can be calculated as follows:

$$\mathbf{X}_l^{(i)'} = \mathbf{X}_l^{(i)} \otimes \text{GELU}(\mathcal{W}_l^{(i)} \odot (\alpha \cdot \text{MaxPool}(\mathbf{X}_l^{(i)}) + \beta \cdot \mathbf{v}_l^{(i)})), \quad (9)$$

where “ $\odot$ ” represents the Hadamard product. This formulation ensures an adaptive balance between local and global information, enhancing the representation of high-frequency details while suppressing noise.

Frequency-Guided Perturbation (FGP). The abundance of speckle noise in SAR images leads to random intensity variations caused by the coherence among small scatterers. This noise generates texture interference within the image, obscuring the boundary between the target and the background, making it easy to misinterpret background noise as significant features, thereby compromising the precision of target detection.

As shown in Figure 5, the FGP module standardizes the feature map and adjusts the feature mean and variance to a unified scale.

$$\begin{aligned} \mu_{l,\text{ori}}^{(i)} &= \frac{1}{H_l W_l} \sum_{p=1}^{H_l} \sum_{q=1}^{W_l} \mathbf{A}_{l,pq}^{(i)} \\ \sigma_{l,\text{ori}}^{(i)} &= \sqrt{\frac{1}{H_l W_l} \sum_{p=1}^{H_l} \sum_{q=1}^{W_l} \left(\mathbf{A}_{l,pq}^{(i)} - \mu_{l,\text{ori}}^{(i)} \right)^2}, \\ \mathbf{A}_{l,\text{normed}}^{(i)} &= \frac{\mathbf{A}_l^{(i)} - \mu_{l,\text{ori}}^{(i)}}{\sigma_{l,\text{ori}}^{2,(i)} + \epsilon}. \end{aligned} \quad (10)$$

where $\mu_{l,ori}^{(i)}$, $\sigma_{l,ori}^{(i)}$, and $\mathbf{A}_{l,normed}^{(i)}$ denote $\mathbf{A}_l^{(i)}$, meaning the results of the mean, standard deviation, and standardization, respectively. Assuming the size of the training batch is B, we randomly rearrange the features $\mathbf{A}_l^{(i,b)} \in \mathbb{R}^{B \times \frac{H}{2^l} \times \frac{W}{2^l} \times C_l}$ within the batch. Specifically, we firstly randomly shuffle the sample order within the batch to obtain a new permutation index perm:

$$\mathbf{A}_{l,perm}^{(i,b)} = RandPerm(\mathbf{A}_l^{(i,b)}), \quad (11)$$

Here, $\mathbf{A}_{l,perm}^{(i,b)}$ represents the permuted features with an average of $\mu_{l,perm}^{(i)}$ and a standard deviation of $\sigma_{l,perm}^{(i)}$, indicating the modified feature distribution. This random modification obscures background information, compelling the model to depend on global characteristics instead of local signals to identify the target. Subsequently, we employ weighted adjustment to merge the mean and standard deviation of the original with those of the permuted features, resulting in the adjusted mean $\mu_{l,adjusted}^{(i)}$ and the adjusted standard deviation $\sigma_{l,adjusted}^{(i)}$ as demonstrated:

$$\begin{aligned} \mu_{l,adjusted}^{(i)} &= \gamma_l^{(i)} \cdot \mu_{l,ori}^{(i)} + (1 - \gamma_l^{(i)}) \cdot \mu_{l,perm}^{(i)}, \\ \sigma_{l,adjusted}^{(i)} &= \gamma_l^{(i)} \cdot \sigma_{l,ori}^{(i)} + (1 - \gamma_l^{(i)}) \cdot \sigma_{l,perm}^{(i)}. \end{aligned} \quad (12)$$

$$\gamma_l^{(i)} = 1 - \lambda_l^{(i)}, \quad (13)$$

Here, $\gamma_l^{(i)} \in \mathbb{R}^{1 \times 1 \times C_l}$ signifies the perturbation coefficient, a parameter that dictates the level of disturbance applied. This coefficient $\gamma_l^{(i)}$ is influenced by the weight in the frequency domain, denoted as $\lambda_l^{(i)}$, after which characteristic disturbances are adaptively adjusted. A large value of $\lambda_l^{(i)}$ signifies that high-frequency information within the image is crucial and heavily weighted for the detection task. To prevent excessive alterations to these details, it is essential to diminish the perturbation coefficient. Conversely, if the detailed information presently has a minor role in the task, or if the background information is more intricate, enhancing the perturbation coefficient can help mitigate the impact of background noise. The resulting perturbation features are represented as follows:

$$\mathbf{A}_l^{(i)'} = \mathbf{A}_{l,normed}^{(i)} \otimes \sigma_{l,adjusted}^{(i)} + \mu_{l,adjusted}^{(i)}. \quad (14)$$

It combines $\mathbf{A}_{l,normed}^{(i)}$ with $\mu_{l,adjusted}^{(i)}$ and $\sigma_{l,adjusted}^{(i)}$ to ensure that the output feature map not only retains the necessary high-frequency details but also focuses more on global features when the background noise is complex. This dynamic adjustment mechanism enables the FGP module to deal with the dependence of the target on background information in different scenarios.

The final output of the AFMP module is

$$\mathbf{Z}_l^{(i)} = \mathbf{A}_l^{(i)'} + \mathbf{X}_l^{(i)'}. \quad (15)$$

3.4. Loss Function

The final loss function is composed of a weighted binary cross-entropy loss $\mathcal{L}_{BCE}^w$, a weighted CIoU loss $\mathcal{L}_{CIoU}^w$, and a distribution focal loss $\mathcal{L}_{DFL}^w$, where each component has a specific weight to ensure an appropriate influence on the total loss. The resultant loss function can be formulated as

$$\mathcal{L} = \lambda_{cls} \mathcal{L}_{BCE}^w + \lambda_{iou} \mathcal{L}_{CIoU}^w + \lambda_{dfl} \mathcal{L}_{DFL}^w. \quad (16)$$

4. Experiments

4.1. Dataset and Experimental Setup

(1) *The LS-SSDD-v1.0* [47] is constructed from Sentinel-1 SAR data released by the European Space Agency. The dataset is composed of Sentinel-1 acquisitions in the interferometric wide swath imaging mode. It contains two polarization channels, VV and the VH, which are the default polarimetric combinations over maritime surveillance areas. Following the official setting, large-scene images are cut into 9000 sub-images of 800×800 pixels, where ships are usually small and thus challenging to detect. According to the official division, 6000 sub-images are used for training and 3000 for testing.

(2) *SSDD* [48] integrates SAR images collected from three sensors, namely RadarSat-2, TerraSAR-X, and Sentinel-1. It includes multiple polarization modes, namely HH, HV, VV, and VH. The dataset contains 1160 SAR images of 500×500 pixels with 2358 ship instances, and the spatial resolution spans from 1 m to 15 m. The official split uses 928 images for training and the remaining images for testing.

(3) *SAR-Ship-Dataset* [49] was built from Gaofen-3 and Sentinel-1 data and contains 43,819 ship chips of 256×256 pixels covering 59,535 ship instances. The dataset involves multiple SAR acquisition modes, including ultrafine strip-map, fine strip-map 1, fine strip-map 2, quad-polarization strip-map 1, S3 strip-map, and interferometric wide-swath. It provides samples with multiple spatial resolutions, including 3 m, 5 m, 8 m, 10 m, and 25 m, and it supports single-polarization, dual-polarization, and full-polarization settings. Since no official split has been released, we randomly split the dataset into training and testing sets at a ratio of 8 to 2 following common practice in comparative methods.

(4) *AIR-SARShip-2.0* [50] was collected from Gaofen-3 high-resolution SAR imagery. According to the dataset description, Gaofen-3 data are acquired under strip-map imaging mode and spotlight imaging mode, and the dataset provides meter-level resolutions, including 1 m and 3 m, with single-polarization measurements. It contains 300 SAR images of 1000×3000 pixels covering various complex scenes, including ports, islands, reefs, and open sea surfaces. Since no official split standard is provided, the dataset was randomly divided into training and testing sets at a ratio of 8 to 2 following the settings in other methods, which is suitable for high-resolution SAR ship detection under complex backgrounds.

We performed experiments on four NVIDIA GeForce GTX3090 Gpus with 24 GB of memory with Pytorch-1.12.1 and python 3.9, with the batch size set to 96. The size of the training images was 640 for the LS-SSDD-v1.0, SSDD, and AIR-SARShip-2.0 datasets and 256 for the SAR-Ship-Dataset. We trained the model 300 times in total, and the Adam optimizer was used with an initial learning rate of 0.001, and the weight decay and momentum were set to 5×10^{-4} and 0.9. To ensure a fair comparison, all comparative methods were trained and evaluated under the same protocol on each dataset, including the same data splits, input resolution, batch size, training epochs, and the same data augmentation pipeline (random flipping, random scaling, and mosaic augmentation).

4.2. Evaluation Metrics

In order to systematically evaluate the performance of the model in the ship detection tasks on SAR images, this study used the precision, recall and F1 score, along with the mean average precision (mAP), as primary evaluation metrics. mAP is widely used in object detection tasks, as it comprehensively measures the performance of a model in terms of precision and recall, reflecting the stability of the model across different detection thresholds. mAP calculates the average precision (AP) for each class by taking the area under the precision–recall curve. The metrics are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (17)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (18)$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (19)$$

where TP denotes the number of correctly detected targets, FP represents regions that are incorrectly detected as targets, and FN represents the number of missed targets. AP is calculated as follows:

$$\text{AP} = \frac{1}{n} \sum_{r \in \{\frac{1}{n}, \frac{2}{n}, \dots, 1\}} P(r), \quad (20)$$

where n represents the number of sampling points on the recall curve used to calculate AP, and $P(r)$ represents the precision of a recall of r . The mAP is defined as the average of the AP values in all categories, as follows:

$$\text{mAP} = \frac{1}{N} \sum_{k=1}^N \text{AP}_k, \quad (21)$$

where N denotes the total number of target categories, and AP_k represents the average precision for the k -th category. To comprehensively assess the detection performance of the model, AP is calculated under various IoU thresholds, including AP_{50} (IoU = 0.5), AP_{75} (IoU = 0.75), and AP_{50-95} , which is the mean AP across IoU thresholds from 0.5 to 0.95 with a step size of 0.05. In this study, since the detection target is limited to the category of the ship, the calculation of mAP simplifies to the AP value for this single class.

4.3. Comparison with State-of-the-Arts Methods

(1) Quantitative evaluation: Tables 1–4 present the detailed comparison results of our proposed FMP-Net against 18 other state-of-the-art methods in multiple benchmark datasets. These methods include the following: one-stage object detection networks: Atss [51], CenterNet [52], Fcos [53], Gfl [54], Retinanet [55], YOLOv3n-tiny [10], YOLOv5n [11], YOLOv6n [12], YOLOv9t [13], YOLOv8n [14], YOLOv10n [15], and YOLOv11n [16]; two-stage object detection networks: Cascade-RCNN [56], Faster-RCNN [9], Grid-RCNN [57], and Libra-RCNN [58]; and end-to-end networks: Dino [59] and RTDetResNet50 [60]. It can be seen that our method significantly outperforms these comparison methods on all benchmark datasets and achieves higher detection performance. Among them, our proposed FMP-Net achieves the best or second-best performance in several evaluation metrics, demonstrating excellent advantages. In particular, our method maintains a lightweight architecture with only 2.78M parameters and 7.8G FLOPs, which are significantly lower metrics than those of most comparison methods.

On the LS-SSDD-v1.0 dataset. As shown in Table 1, we performed a detailed analysis on inshore and offshore scenarios, which represent different levels of detection difficulty. In inshore scenarios, FMP-Net achieves a precision of 67.7%, a recall of 51.1%, and an F1 score of 58.2%, significantly outperforming other methods. Compared to the second-best, YOLOv8s, the F1 score is improved by 2.1% (from 56.1% to 58.2%), and AP50 is increased by 4.2% (from 52.8% to 57.0%). This superior performance is attributed to the Frequency-aware Modulation and Perturbation strategy, which effectively distinguishes ships from strongly scattering harbor facilities and coastal structures. In offshore scenarios, AP75 is significantly higher than the second-best, YOLOv5n (an increase of 1.8%), where AP50 and mAP reach 93.1% and 40.6%, respectively. This indicates that the method has a strong detection ability in offshore background. It should be noted that while most methods show better performance in offshore scenarios than inshore scenarios, the proposed FMP-Net has a smaller performance gap (36.5% difference in AP50) between these two scenarios than

other methods (for example, YOLOv6n shows a 47.3% gap). This shows that the proposed method maintains a more consistent performance under different environmental conditions.

Table 1. Quantitative comparison of FMP-Net with 18 SOTA methods for SAR ship detection on the LS-SSDD-v1.0 dataset. **Bold** indicates the best performance, and underline indicates the second-best performance.

Methods	All						Inshore						Offshore						Params (M)	FLOPs (G)
	P	R	F1	AP50	mAP	AP75	P	R	F1	AP50	mAP	AP75	P	R	F1	AP50	mAP	AP75		
Atss [51]	84.2	66.0	74.0	73.3	28.9	13.5	60.5	34.8	44.2	39.5	13.5	5.2	88.3	86.6	87.4	89.9	36.8	18.1	32.11	80.5
Cascade-RCNN [56]	83.5	62.9	71.8	73.9	30.5	17.6	59.1	30.9	40.6	44.5	18.2	<u>10.2</u>	86.7	83.8	85.2	88.6	36.5	18.0	69.15	118.7
CenterNet [52]	82.8	69.9	75.8	71.1	25.8	10.4	61.1	41.3	49.3	39.9	12.9	4.0	89.1	87.7	88.4	86.5	32.8	14.6	32.11	78.6
Dino [59]	82.0	67.0	74.0	73.6	27.8	11.1	63.3	41.2	49.9	44.6	15.1	5.2	86.4	84.5	85.5	88.1	34.8	15.2	47.54	118.8
Faster-RCNN [9]	87.9	64.7	74.5	76.1	31.4	17.5	68.0	32.4	43.9	48.3	<u>20.2</u>	11.1	89.9	86.3	88.1	90.4	37.1	20.6	41.35	90.9
Fcos [53]	84.4	66.8	74.6	72.4	27.1	11.3	57.9	37.4	45.4	37.9	12.4	4.8	89.2	<u>87.8</u>	88.5	89.2	34.9	15.7	32.11	78.6
Gfl [54]	85.6	62.2	72.1	69.9	28.1	13.5	58.2	30.4	40.0	35.1	13.0	5.8	89.0	84.9	86.9	89.2	36.5	17.9	32.26	81.7
Grid-RCNN [57]	82.6	63.1	71.5	72.4	29.5	15.9	54.0	34.3	42.0	41.0	16.2	9.6	90.0	80.9	85.2	22.3	36.2	19.4	-	-
Libra-RCNN [58]	79.7	63.2	70.5	71.5	29.2	15.4	52.4	33.7	41.0	37.5	14.3	7.9	87.0	82.3	84.6	88.5	36.6	19.2	41.61	92.6
Retinanet [54]	78.0	61.7	68.9	67.0	24.5	9.1	45.0	29.1	35.3	27.4	9.1	3.2	87.4	85.3	86.3	87.9	33.3	13.3	36.33	81.6
YOLOv3-t [10]	81.3	63.3	71.2	73.0	30.3	17.3	54.2	37.2	44.1	40.9	15.5	7.9	87.7	81.8	84.7	88.9	38.0	22.5	12.13	18.9
RTDetr-ResNet50 [60]	69.9	58.2	38.8	58.2	25.3	12.1	45.8	33.6	38.8	31.3	11.7	5.1	75.4	75.5	75.4	82.5	33.2	16.2	-	-
YOLOv5n [11]	<u>85.9</u>	68.4	76.2	77.0	31.9	17.7	60.7	42.6	50.1	47.3	18.0	8.6	90.6	87.5	<u>89.0</u>	<u>92.7</u>	<u>39.6</u>	<u>23.1</u>	2.50	<u>7.1</u>
YOLOv6n [12]	83.5	68.9	73.7	76.4	30.7	15.4	56.5	42.4	48.4	43.7	15.0	6.5	89.6	85.3	87.4	91.0	38.2	20.6	4.23	11.8
YOLOv9t [13]	85.5	68.2	75.8	77.6	31.6	16.6	61.9	45.3	52.3	50.0	18.4	7.7	89.1	87.4	88.2	92.1	38.7	22.1	1.97	7.6
YOLOv9s [13]	83.1	<u>72.1</u>	<u>77.2</u>	<u>79.0</u>	<u>32.5</u>	<u>18.1</u>	66.1	47.3	55.2	52.5	19.6	8.6	<u>91.0</u>	86.4	88.6	92.5	39.5	23.0	7.2	27.4
YOLOv8s [14]	84.1	67.7	75.0	75.6	29.8	14.5	64.7	49.2	52.5	49.2	17.5	6.2	88.8	86.1	87.4	89.6	36.6	19.6	3.01	8.1
YOLOv8n [14]	83.8	69.9	76.3	78.7	31.7	16.6	66.1	48.7	<u>56.1</u>	<u>52.8</u>	19.3	8.5	88.5	87.3	86.2	91.6	38.5	22.1	11.13	28.4
YOLOv10n [15]	80.7	67.8	73.7	76.7	31.5	16.8	62.9	46.3	53.4	51.2	19.6	9.6	85.2	83.5	84.4	90.2	38.0	21.3	2.27	8.2
YOLOv10s [15]	80.7	67.7	73.7	75.7	30.8	16.4	61.5	44.0	51.7	49.4	18.1	8.1	86.4	83.1	84.7	89.8	37.8	21.3	8.1	24.8
YOLOv11n [16]	84.7	70.1	76.7	78.1	31.5	16.6	63.6	47.2	54.2	51.5	18.8	8.2	89.2	87.2	88.2	91.9	38.7	22.1	2.58	6.3
YOLOv11s [16]	87.4	67.3	76.0	78.0	32.4	17.8	64.6	44.6	52.8	49.0	18.8	8.5	90.9	87.6	89.2	93.3	39.9	23.7	9.4	21.5
Ours	85.9	72.8	78.8	81.0	33.6	18.1	<u>67.7</u>	51.1	58.2	57.0	21.0	7.7	91.2	88.2	89.7	93.1	40.6	24.9	2.78	7.8

Table 2. Quantitative comparison of FMP-Net with 18 SOTA methods for SAR ship detection on the SSDD dataset. **Bold** indicates the best performance, and underline indicates the second-best performance.

Methods	P	R	F1	AP50	mAP	AP75
Atss [51]	93.5	87.2	90.2	93.8	65.0	78.1
Cascade-RCNN [56]	95.6	85.2	90.1	94.5	<u>66.7</u>	74.1
CenterNet [52]	90.7	84.8	87.6	90.0	52.9	56.5
Dino [59]	93.0	82.1	87.2	91.7	62.2	73.1
Faster-RCNN [9]	93.3	81.1	86.8	91.3	46.8	43.2
Fcos [53]	95.1	87.7	91.3	93.5	55.7	61.9
Gfl [54]	91.7	83.2	87.3	90.3	55.0	56.0
Grid-RCNN [57]	95.1	85.0	89.7	92.6	56.0	56.0
Libra-RCNN [58]	92.2	88.8	90.5	94.9	64.8	78.2
Retinanet [54]	82.7	82.1	82.4	87.2	42.3	35.9
YOLOv3-t [10]	95.5	89.1	92.2	95.3	63.5	75.6
RTDetr-ResNet50 [60]	91.0	88.9	89.9	95.0	61.7	73.5
YOLOv5n [11]	95.2	95.9	95.2	97.7	66.2	78.1
YOLOv6n [12]	95.5	94.4	95.2	97.2	64.4	73.9
YOLOv9t [13]	<u>97.6</u>	93.3	95.2	97.0	66.7	77.9
YOLOv9s [13]	95.9	96.0	95.8	97.2	66.0	77.3
YOLOv8n [14]	94.6	94.2	94.8	97.1	65.6	75.5
YOLOv8s [14]	96.8	96.8	<u>95.8</u>	97.7	66.2	77.8
YOLOv10n [15]	94.9	93.4	94.5	<u>97.8</u>	65.4	77.8
YOLOv10s [15]	92.7	95.4	94.0	97.4	65.0	76.0
YOLOv11n [16]	95.6	93.9	95.1	97.5	65.5	79.5
YOLOv11s [16]	95.3	95.0	95.1	96.9	66.0	77.1
Ours	97.7	<u>96.2</u>	97.0	98.3	67.4	<u>79.3</u>

On the SSDD dataset. As shown in Table 2, our method achieves an AP50 of 98.3% and mAP of 67.4%, outperforming all comparison methods. Specifically, compared with the second-best method, YOLOv8s (mAP: 66.2%; AP75: 77.8%), our mAP and AP75 are improved by 1.2% and 1.5%, respectively. The F1 value reaches 97.0%, which is 1.2% higher than that of YOLOv8s. Additionally, our precision reaches 97.7% and recall reaches 96.2%, demonstrating the superior and balanced detection capabilities of FMP-Net.

Table 3. Quantitative comparison of FMP-Net with 18 SOTA methods for SAR ship detection on the SAR-Ship-Dataset. **Bold** indicates the best performance, and underline indicates the second-best performance.

Methods	P	R	F1	AP50	mAP	AP75
Atss [51]	95.2	93.9	94.6	96.9	67.1	79.2
Cascade-RCNN [56]	89.4	89.6	89.5	93.7	57.7	62.9
CenterNet [52]	94.3	93.8	94.0	96.7	64.2	74.6
Dino [59]	96.5	95.2	95.8	96.5	67.0	80.7
Faster-RCNN [9]	92.5	88.9	90.7	94.4	57.4	63.4
Fcos [53]	95.7	<u>95.0</u>	95.3	<u>97.5</u>	<u>67.8</u>	80.6
Gfl [54]	95.7	<u>94.5</u>	95.1	97.3	67.8	80.2
Grid-RCNN [57]	91.0	87.4	89.1	84.0	56.5	62.6
Libra-RCNN [58]	94.6	93.3	93.9	96.7	67.1	<u>80.6</u>
Retinanet [54]	92.9	93.1	93.0	96.3	64.5	74.5
YOLOv3-t [10]	91.8	86.5	89.1	93.6	56.5	61.4
RTDetr-ResNet50 [60]	92.6	93.2	92.9	95.5	59.9	67.2
YOLOv5n [11]	90.8	90.4	90.6	94.8	58.5	64.7
YOLOv6n [12]	93.8	91.6	92.7	95.6	61.5	70.6
YOLOv9t [13]	91.2	90.6	90.9	94.8	57.9	64.3
YOLOv9s [13]	92.1	91.4	91.7	95.7	61.0	69.4
YOLOv8n [14]	93.0	92.4	92.7	96.4	62.4	71.9
YOLOv8s [14]	94.6	93.7	94.1	97.3	67.3	80.0
YOLOv10n [15]	91.2	88.9	90.5	94.6	57.5	63.4
YOLOv10s [15]	90.4	91.0	90.7	94.8	58.9	66.1
YOLOv11n [16]	90.9	91.3	91.1	95.1	60.3	68.9
YOLOv11s [16]	92.1	92.7	92.4	96.3	63.3	73.2
Ours	<u>96.0</u>	94.6	<u>95.3</u>	97.8	68.9	82.0

Table 4. Quantitative comparison of FMP-Net with 18 SOTA methods for SAR ship detection on the AIR-SARShip-2.0 dataset. **Bold** indicates the best performance, and underline indicates the second-best performance.

Methods	P	R	F1	AP50	mAP	AP75
Atss [51]	68.5	59.9	63.9	64.5	27.4	17.9
Cascade-RCNN [56]	71.2	71.0	71.1	73.2	32.5	32.5
CenterNet [52]	48.2	44.3	46.2	41.6	13.3	3.5
Dino [59]	54.6	60.8	53.8	53.8	26.1	22.1
Faster-RCNN [9]	71.0	66.4	68.6	72.3	35.9	29.0
Fcos [53]	57.1	60.6	58.8	59.0	22.2	10.2
Gfl [54]	59.3	57.4	58.3	58.1	25.5	14.6
Grid-RCNN [57]	68.3	68.2	68.2	68.5	30.2	20.5
Libra-RCNN [58]	60.1	59.7	59.9	59.0	24.0	12.4
Retinanet [54]	59.2	57.8	58.5	57.7	20.6	6.5
YOLOv3-t [10]	79.4	68.2	74.0	75.5	75.5	38.9
RTDetr-ResNet50 [60]	63.0	67.9	65.4	66.2	32.9	27.1
YOLOv5n [11]	<u>80.9</u>	77.6	79.2	<u>81.8</u>	45.0	47.4
YOLOv6n [12]	74.3	71.2	72.8	75.2	38.9	34.3
YOLOv9t [13]	78.6	79.5	79.1	81.7	45.8	42.7
YOLOv9s [13]	81.9	72.8	77.1	80.2	43.6	41.2
YOLOv8n [14]	78.1	77.1	77.6	80.2	46.2	46.8
YOLOv8s [14]	76.3	72.4	74.3	80.3	42.9	41.0
YOLOv10n [15]	78.5	<u>80.9</u>	79.6	81.0	46.8	<u>48.6</u>
YOLOv10s [15]	67.6	69.8	68.7	72.6	37.4	33.0
YOLOv11n [16]	77.7	80.9	79.0	80.8	<u>47.6</u>	47.1
YOLOv11s [16]	76.1	77.6	76.8	80.5	44.2	43.9
Ours	79.5	79.4	<u>79.4</u>	83.0	48.5	51.9

On the SAR-Ship-Dataset dataset. As shown in Table 3, for AP75, our FMP-Net achieves 82.0%, exceeding the second-best Dino (80.7%) by 1.3%. The strong AP75 score

indicates our model's exceptional ability in precise target localization. In terms of mAP, which evaluates detection performance across different IoU thresholds, our method reaches 68.9%, surpassing Dino (67.0%) by 1.9%. The F1 score reaches 95.3% with 96.0% precision and 94.6% recall, indicating balanced detection performance.

On the AIR-SARShip-2.0 dataset. As shown in Table 4, our FMP-Net achieves superior performance across all evaluation metrics. For AP75, our method reaches 51.9%, notably outperforming the second-best YOLOv10n (48.6%) by 3.3%. This significant improvement in AP75 demonstrates our model's enhanced capability in accurate target localization. In terms of AP50, our approach attains 83.0%, surpassing the second-best YOLOv5n (81.8%) by 1.2%. For mAP, which evaluates detection performance across various IoU thresholds, our method achieves 48.5%, exceeding YOLOv10n (46.8%) by 1.7%. The remarkable improvement in AP75 (3.3% higher than that of the second-best method) highlights the effectiveness of our frequency-aware modulation and perturbation strategy in accurately localizing targets, even in complex scenarios like ports, islands, and reefs.

(2) Qualitative evaluation: In Figure 7, we show the qualitative comparison of our method to several representative SOTA methods in some challenging cases, with correctly detected targets, missed targets, and false alarms marked blue, red, and yellow, respectively. Our method FMP-Net achieves more competitive visual performance mainly in the following aspects: Regarding small objects in sea clutter interference (first row), traditional detection methods often struggle in this scene due to the high background noise level, and our method has better prediction and localization ability for small objects. In the complex inshore scene (second, third, and fourth rows), the traditional model is prone to false detections, while our method suppresses irrelevant interference and maintains the accuracy in recognizing the true target. In strongly scattering, side-lobe interference scenarios (fifth and sixth rows), which usually lead to false detections due to stacked reflections, our method achieves robust performance even in highly scattering environments. More accurate detection of the target is realized in the low-contrast scene (seventh row). These qualitative results demonstrate the effectiveness of our approach in a variety of challenging SAR image scenarios.

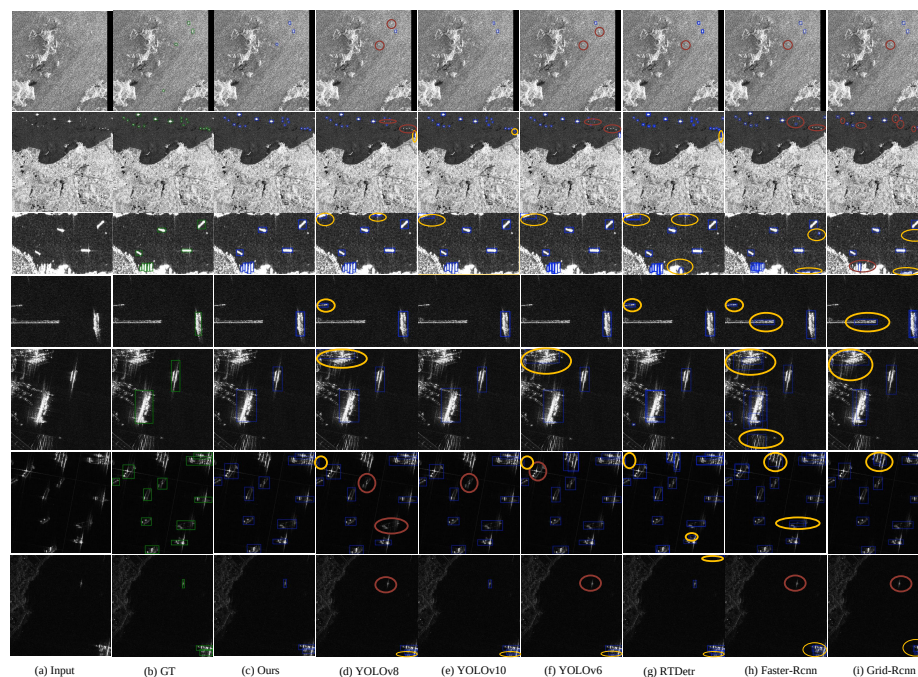


Figure 7. Visual comparison of our method with six representative state-of-the-art methods. Our method is capable of tackling challenging cases (e.g., small objects under sea clutter (first row), complex inshore scene (second, third, and fourth rows), strong scattering (fifth and sixth rows), and low-contrast scene (seventh row)).

4.4. Ablation Study

This section delves into the ablation study conducted on the LS-SSDD-v1.0 and SSDD datasets. Note that the three proposed techniques are not independent plug-ins. Specifically, the variance modulation is defined on the directional high-frequency weights extracted by AWFA, and the FGP module was designed for Frequency-Guided Perturbation that relies on the AWFA-derived frequency weights.

(1) *The effect of the modulation strategy.* As shown in Table 5, on LS-SSDD-v1.0, the AP50 and AP75 of the model also increased from 75.6% and 14.5% to 77.7% and 17.8% after adding AWFA, respectively. After adding variance modulation to the basis of the AWFA module, the AP50 and AP75 of the model improve to 78.8% and 18.3%, respectively. Similarly, on SSDD data, AP75 improved from 75.5% to 76.7% after adding the AWFA module and then reached 78.5% after adding variance modulation. This indicates the synergistic effect of the AWFA module and modulation. Firstly, the AWFA module enhances the high-frequency features in the SAR images, making the model more sensitive to boundaries and details. The variance modulation further dynamically adjusts the high-frequency features so that the model can extract effective information from complex backgrounds and suppress irrelevant noise.

Table 5. The effect of the modulation strategy.

AWFA	Modulation	LS-SSDD-v1.0						SSDD						Params (M)	FLOPs (G)
		P	R	F1	AP50	mAP	AP75	P	R	F1	AP50	mAP	AP75		
		84.1	67.7	75.0	75.6	29.8	14.5	94.6	94.2	94.8	97.1	65.6	75.5	3.01	8.1
✓		84.6	67.2	74.9	77.7	32.0	17.8	96.6	93.2	94.9	97.7	65.7	76.7	3.00	7.9
✓	✓	85.7	70.7	77.5	78.8	32.6	18.3	95.3	95.7	95.5	97.8	66.3	78.5	3.10	8.0

(2) *The effect of the FGP module.* As shown in Table 6, on LS-SSDD-v1.0, F1 and AP50 improved from 77.0% and 78.4 to 78.3% and 80.2% after adding the FGP module, respectively. Similarly, on SSDD, F1 and AP75 improved from 95.1% and 78.2% to 96.0% and 79.0%, respectively. This indicates that the FGP module forces the model to pay more attention to the essential characteristics of the target and weakens the influence of background information during the learning process by imposing dynamic disturbances on the feature space.

Table 6. The effect of the FGP module.

FGP Status	LS-SSDD-v1.0						SSDD						Param (M)	FLOPs (G)
	P	R	F1	AP50	mAP	AP75	P	R	F1	AP50	mAP	AP75		
w/o FGP	84.3	70.8	77.0	78.4	32.4	18.0	96.0	97.6	95.1	97.6	66.4	78.2	3.10	8.0
w/ FGP	84.4	73.0	78.3	80.2	33.2	18.6	97.0	94.9	96.0	97.9	67.2	79.0	3.10	8.0

(3) *The strategy for selecting the interference coefficient γ in the FGP module.* As shown in Table 7, we explore the effect of different values of the interference coefficients γ on the performance of the model. In the experiment, three schemes were compared: no disturbance, random disturbance coefficient selection, and dynamic adjustment of the disturbance coefficient according to the frequency weight. After using random values such as γ , in the LS dataset, although AP50 increased slightly, AP75 decreased by 0.7%. This indicates that random perturbations introduce some uncertainty, and inappropriate perturbations can degrade the performance of the model. When $\gamma = 1 - \lambda$, compared to randomly selected γ in LS-SSDD-v1.0, AP50 improves by 1.5%, AP75 improves by 1.3%, and F1 improves by 1.0% in SSDD. The experimental results show that the dynamic adjustment strategy based on frequency weight has more advantages than random disturbance, so that the model can weaken the interference of irrelevant background while maintaining the attention to the important features of the target.

Table 7. The strategy for selecting the interference coefficient γ in the FGP module.

γ	LS-SSDD-v1.0						SSDD					
	P	R	F1	AP50	mAP	AP75	P	R	F1	AP50	mAP	AP75
base	84.3	70.8	77.0	78.4	32.4	18.0	96.0	97.6	95.1	97.6	66.4	78.2
random	85.7	70.7	77.6	78.7	32.3	17.3	94.6	94.6	95.0	97.4	66.6	79.0
$1 - \lambda$	84.4	73.0	78.3	80.2	33.2	18.6	97.0	94.9	96.0	97.9	67.2	79.0

(4) *Effectiveness analysis of the AWFDwn and MS-FMP module.* As shown in Table 8, adding only the AWFDwn module increases the mAP model in LS-SSDD-v1.0 by 2.9% and AP75 by 3.0% compared to the baseline. In SSDD, AP75 improves by 3.1%. When only the MS-FMP module is introduced, AP75 and F1 improve by 3.4% and 3.3%, respectively, in LS-SSDD-v1.0, and by 3.5% and 2.2% in SSDD. When the two modules are combined, the model achieves the best detection performance. Compared to the baseline, the F1, AP50, and mAP in LS-SSDD-v1.0 increase to 78.8%, 81.0%, and 33.6%, respectively. In SSDD, the F1, AP50, and mAP increase to 97.0%, 98.3%, and 67.4%, respectively. The AWFDwn module effectively reduces the loss of high-frequency details during feature downsampling by preserving high-frequency information in the image. Based on this, MS-FMP further enhances these high-frequency features and effectively suppresses the interference of background noise, thereby improving the SAR target detection ability of the model in complex backgrounds.

Table 8. Effectiveness analysis of AWFDwn and MS-FMP modules.

AWFDwn	MS-FMP	LS-SSDD-v1.0						SSDD						Params (M)	FLOPs (G)
		P	R	F1	AP50	mAP	AP75	P	R	F1	AP50	mAP	AP75		
		84.1	67.7	75.0	75.6	29.8	14.5	94.6	94.2	94.8	97.1	65.6	75.5	3.01	8.1
✓		84.2	70.2	76.6	77.8	32.7	17.5	95.6	93.8	95.0	97.4	66.0	78.6	2.69	7.8
	✓	84.4	73.0	78.3	80.2	33.2	18.6	97.0	94.9	96.0	97.9	67.2	79.0	3.10	8.0
✓	✓	85.9	72.8	78.8	81.0	33.6	18.1	97.7	96.2	97.0	98.3	67.4	79.3	2.78	7.8

(5) *Effectiveness comparison of FPM stage in different object detection networks.* The FPM stage improves the detection accuracy across different types of networks, including one-stage models (e.g., CenterNet, YOLOv5, YOLOv6, YOLOv9, YOLOv10), two-stage models (e.g., Faster-RCNN, Grid-RCNN, and Librar-RCNN), and end-to-end models (e.g., RTDetr-ResNet50 and Dino). For example, as shown in Table 9 on LS-SSDD-v1.0, integrating the FPM stage into Faster-RCNN improves AP50 by 1.8% and mAP by 2.0%. For RTDetr-ResNet50, AP50 is improved by about 10.1%. Integrating the FPM stage into YOLOv5n improves AP50 by 2.8%. These improvements show that the FPM stage, as an effective plug-and-play module, can significantly enhance high-frequency feature extraction capabilities across different detection architectures, and is an important component in improving the robustness and accuracy of detection in complex environments.

4.5. Visualization of Feature Activation Maps

To better understand the effectiveness of the proposed method, we visualize the feature activation maps in Figure 8, where brighter colors indicate stronger activation responses. We selected four representative challenging scenarios for analysis. As shown in the first row, with strong coastal interference, our method produces a more focused activation for ship targets while effectively suppressing coastal background interference, while the baseline shows a distraction of background structure. In the second row of the offshore scene with sea clutter interference, our method produces a more focused and stronger activation for the ship target, while the baseline shows a relatively weak and ambiguous response due to the influence of sea clutter. For the multi-ship case of different sizes (third row),

the baseline exhibits relatively weak and ambiguous responses, and our method maintains consistent and precise activation patterns regardless of the target scale, which demonstrates the effectiveness of our method. In a complex scene with strong scattering interference (fourth row), our method generates clearer and sharper activation maps for each target, indicating excellent discrimination in challenging environments. These visualizations provide intuitive evidence that our Frequency-aware Modulations and Perturbations help the network focus on essential ship features while effectively filtering out background noise and reducing interference.

Table 9. Performance comparison of inserting FPM stage into different object detection networks.

Models	P	R	F1	AP50	mAP	AP75	Params (M)	FLOPs (G)
Two-stage								
Faster-RCNN	87.9	64.7	74.5	76.1	31.4	17.5	41.34	90.898
Faster-RCNN w/FPM Stage	85.8	65.6	74.3	77.9	34.4	21.9	29.70	81.216
Difference	−2.1	+0.9	−0.2	+1.8	+2.0	+3.4		
Grid-RCNN	82.6	63.1	71.5	72.4	29.5	15.9	-	-
Grid-RCNN w/FPM Stage	86.2	63.8	73.3	75.1	32.5	19.7	-	-
Difference	+3.6	+0.7	+1.8	+2.7	+3.0	+3.8		
Libra-RCNN	79.7	63.2	70.5	71.5	29.2	15.4	41.61	92.6
Libra-RCNN w/FPM Stage	83.8	64.5	72.9	76.8	34.7	20.0	29.97	83.0
Difference	+4.1	+1.2	+2.4	+5.3	+5.5	+4.6		
End2end								
RTDetr-ResNet50	69.9	58.2	63.4	64.1	25.4	12.1	41.93	125.6
RTDetr-ResNet50 w/FPM Stage	78.1	66.1	71.6	74.2	28.6	12.2	30.31	106.4
Difference	+8.2	+7.9	+8.2	+10.1	+3.2	+0.1		
Dino	82.0	67.0	74.0	73.6	27.8	11.1	47.54	118.8
Dino w/ FPM Stage	84.1	70.3	76.6	75.6	27.9	11.6	32.36	108.77
Difference	+2.1	+3.3	+2.6	+2.0	+0.1	+0.5		
One-stage								
CenterNet	82.8	69.9	75.8	71.1	25.8	10.4	32.11	78.56
CenterNet w/FPM Stage	83.1	72.0	77.2	73.6	27.1	11.3	20.47	70.13
Difference	+0.3	+2.1	+1.4	+2.5	+1.3	+0.9		
YOLOv5n	85.9	68.4	76.2	77.0	31.9	17.7	2.50	7.1
YOLOv5n w/FPM Stage	85.7	71.2	77.8	79.8	33.0	17.7	2.80	8
Difference	−0.2	+2.8	+1.6	+2.8	+1.1	0.0		
YOLOv6n	83.5	68.9	73.7	76.4	30.7	15.4	4.23	11.8
YOLOv6n w/FPM Stage	82.7	70.3	76.0	78.0	30.8	16.4	3.35	9.6
Difference	−0.8	+1.4	+2.3	+1.6	+0.8	+1.0		
YOLOv9t	85.5	68.2	75.8	77.6	31.6	16.6	1.97	7.6
YOLOv9t w/FPM Stage	85.5	70.0	77.4	80.2	33.1	16.8	1.92	7.6
Difference	0.0	+1.8	+1.6	+2.6	+1.5	+0.9		
YOLOv10n	80.7	68.7	73.7	76.7	31.5	16.8	2.36	8.2
YOLOv10n w/FPM Stage	82.3	69.8	75.5	78.3	32.0	17.7	2.36	6.5
Difference	+1.6	+2.0	+1.8	+1.6	+0.5	+0.9		

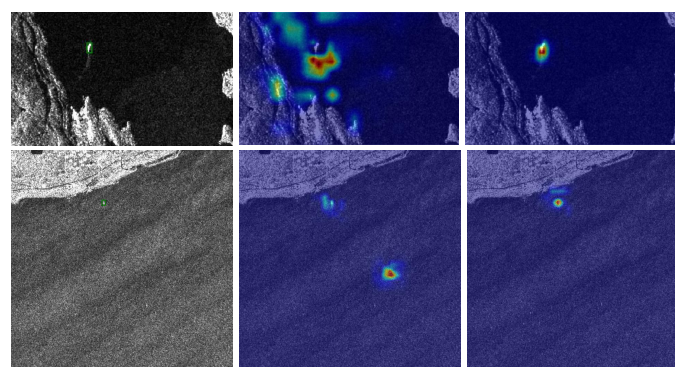


Figure 8. Cont.

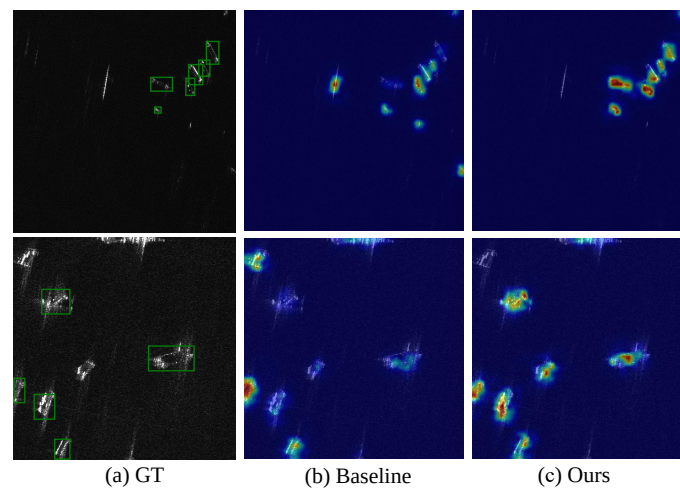


Figure 8. Feature activation map visualization comparison. (a) Ground truth (GT). (b) Baseline. (c) Ours. The rows represent different challenging scenarios: coastal interference, sea clutter interference, multiscale ships, and complex scattering scenes.

5. Conclusions

This paper introduces FMP-Net, a novel Frequency-aware Modulation and Perturbation network for detecting SAR ships, employing a strategy that enhances high-frequency features and reduces background noise. The AWFA module targets high-frequency data in diverse directions, while variance modulation fine-tunes feature description to minimize false detections in complex scenes. The FGP module, guided by frequency domain weights, concentrates on crucial target features while minimizing background effects. Additionally, the AWFDn module maintains high-frequency information during downsampling to retain critical details. These innovations allow FMP-Net to achieve state-of-the-art results in multiple benchmarks with a lightweight design. It excels in both offshore and inshore conditions, especially in challenging circumstances like sea clutter and complex coastlines. The FMP stage shows significant generalization ability, enhancing performance across various detection frameworks. Future research may apply this frequency-aware strategy to other remote sensing tasks and explore its potential in multi-modal fusion.

Author Contributions: Conceptualization and methodology, M.F. and W.Z.; investigation and experiments, M.F. and L.T.; writing—original draft preparation, M.F. and X.Q.; writing—review and editing, W.Z., D.S. and J.Z.; visualization, D.S. and J.Z.; resources and supervision, Y.X. and S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62201467, in part by Innovation Capability Support Program of Shaanxi (Program No. 2024ZC-KJXX-043), and in part by the Young Talent Fund of Xi'an Association for Science and Technology under Grant 959202313088.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: Authors Wencong Zhang, Xiaochen Quan, Dahu Shi were employed by the company Hikrobot Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Wang, R.; Su, T.; Xu, D.; Chen, J.; Liang, Y. MIGA-Net: Multi-view image information learning based on graph attention network for SAR target recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 10779–10792. [CrossRef]
- Geng, J.; Ma, W.; Jiang, W. Causal Intervention and Parameter-Free Reasoning for Few-Shot SAR Target Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 12702–12714. [CrossRef]
- Huang, B.; Zhang, T.; Quan, S.; Wang, W.; Guo, W.; Zhang, Z. Scattering Enhancement and Feature Fusion Network for Aircraft Detection in SAR Images. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *35*, 1936–1950. [CrossRef]
- Huang, S.Q.; Liu, D.Z.; Gao, G.Q.; Guo, X.J. A novel method for speckle noise reduction and ship target detection in SAR images. *Pattern Recognit.* **2009**, *42*, 1533–1542. [CrossRef]
- Guo, H.; Yang, X.; Wang, N.; Gao, X. A CenterNet++ model for ship detection in SAR images. *Pattern Recognit.* **2021**, *112*, 107787. [CrossRef]
- Xing, X.; Ji, K.; Zou, H.; Sun, J.; Zhou, S. High resolution SAR imagery ship detection based on EXS-C-CFAR in Alpha-stable clutters. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; pp. 316–319.
- Leng, X.; Ji, K.; Yang, K.; Zou, H. A bilateral CFAR algorithm for ship detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1536–1540. [CrossRef]
- Wang, C.; Bi, F.; Zhang, W.; Chen, L. An intensity-space domain CFAR method for ship detection in HR SAR images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 529–533. [CrossRef]
- 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.* **2016**, *39*, 1137–1149. [CrossRef] [PubMed]
- Farhadi, A.; Redmon, J. Yolov3: An incremental improvement. In Proceedings of the Computer Vision and Pattern Recognition; Springer: Berlin/Heidelberg, Germany, 2018; Volume 1804, pp. 1–6.
- Jocher, G. YOLOv5 by Ultralytics. 2020. Available online: <https://zenodo.org/records/7347926> (accessed on 7 January 2026).
- Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976. [CrossRef]
- Wang, C.Y.; Yeh, I.H.; Mark Liao, H.Y. Yolov9: Learning what you want to learn using programmable gradient information. In Proceedings of the European Conference on Computer Vision—ECCV 2024, Milan, Italy, 29 September–4 October 2024; pp. 1–21.
- Reis, D.; Kupec, J.; Hong, J.; Daoudi, A. Real-time flying object detection with YOLOv8. *arXiv* **2023**, arXiv:2305.09972. [CrossRef]
- Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. Yolov10: Real-time end-to-end object detection. *arXiv* **2024**, arXiv:2405.14458.
- Jocher, G.; Qiu, J. Ultralytics YOLO11. 2024. Available online: <https://docs.ultralytics.com/models/yolo11/#usage-examples> (accessed on 7 January 2026).
- Fu, J.; Sun, X.; Wang, Z.; Fu, K. An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 1331–1344. [CrossRef]
- Yang, X.; Zhang, X.; Wang, N.; Gao, X. A robust one-stage detector for multiscale ship detection with complex background in massive SAR images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5217712. [CrossRef]
- Hong, Z.; Yang, T.; Tong, X.; Zhang, Y.; Jiang, S.; Zhou, R.; Han, Y.; Wang, J.; Yang, S.; Liu, S. Multi-scale ship detection from SAR and optical imagery via a more accurate YOLOv3. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6083–6101. [CrossRef]
- Hu, Q.; Hu, S.; Liu, S.; Xu, S.; Zhang, Y.D. FINet: A feature interaction network for SAR ship object-level and pixel-level detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5239215. [CrossRef]
- Zha, C.; Min, W.; Han, Q.; Xiong, X.; Wang, Q.; Xiang, H. SAR ship detection based on salience region extraction and multi-branch attention. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *123*, 103489. [CrossRef]
- Yasir, M.; Liu, S.; Mingming, X.; Wan, J.; Pirasteh, S.; Dang, K.B. ShipGeoNet: SAR image-based geometric feature extraction of ships using convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5202613. [CrossRef]
- He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
- Li, Y.; Mao, H.; Girshick, R.; He, K. Exploring plain vision transformer backbones for object detection. In Proceedings of the European Conference on Computer Vision—ECCV 2022, Tel Aviv, Israel, 23–27 October 2022; pp. 280–296.
- Shen, J.; Bai, L.; Zhang, Y.; Momi, M.C.; Quan, S.; Ye, Z. ELLK-Net: An efficient lightweight large kernel network for SAR ship detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5221514. [CrossRef]
- Yasir, M.; Shanwei, L.; Mingming, X.; Jianhua, W.; Nazir, S.; Islam, Q.U.; Dang, K.B. SwinYOLOv7: Robust ship detection in complex synthetic aperture radar images. *Appl. Soft Comput.* **2024**, *160*, 111704. [CrossRef]
- 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 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.

28. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
29. Zhang, T.; Xie, N.; Quan, S.; Wang, W.; Wei, F.; Yu, W. Polarimetric SAR Ship Detection Based on Sub-Look the Decomposition Technology. *IEEE Trans. Radar Syst.* **2026**, *4*, 35–49. [[CrossRef](#)]
30. Wang, J.; Quan, S.; Xing, S.; Li, Y.; Wu, H.; Meng, W. PSO-based fine polarimetric decomposition for ship scattering characterization. *ISPRS J. Photogramm. Remote Sens.* **2025**, *220*, 18–31. [[CrossRef](#)]
31. Zhang, L.; Liu, Y.; Zhao, W.; Wang, X.; Li, G.; He, Y. Frequency-adaptive learning for SAR ship detection in clutter scenes. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5215514. [[CrossRef](#)]
32. Wang, S.; Cai, Z.; Yuan, J. Automatic SAR ship detection based on multifeature fusion network in spatial and frequency domains. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4102111. [[CrossRef](#)]
33. Zhang, Z.; Li, Y. YOLO-FCNET: Enhancing SAR Ship Detection with Fourier Convolution in YOLOv8. In Proceedings of the International Joint Conference on Artificial Intelligence 2024, Jeju, Republic of Korea, 3 August 2024; pp. 195–207.
34. 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.* **2021**, *60*, 5203213. [[CrossRef](#)]
35. Qiao, C.; Shen, F.; Wang, X.; Wang, R.; Cao, F.; Zhao, S.; Li, C. A novel multi-frequency coordinated module for sar ship detection. In Proceedings of the 2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI), Macao, China, 31 October–2 November 2022; pp. 804–811.
36. Wilson, S.; Fischer, T.; Dayoub, F.; Miller, D.; Sünderhauf, N. SAFE: Sensitivity-aware features for out-of-distribution object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 23565–23576.
37. Qi, L.; Dong, P.; Xiong, T.; Xue, H.; Geng, X. DoubleAUG: Single-domain Generalized Object Detector in Urban via Color Perturbation and Dual-style Memory. *ACM Trans. Multimed. Comput. Commun. Appl.* **2024**, *20*, 1–20. [[CrossRef](#)]
38. Li, S.; Ye, M.; Zhu, X.; Zhou, L.; Xiong, L. Source-free object detection by learning to overlook domain style. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 8014–8023.
39. Song, Z.; Gong, X.; Hu, G.; Zhao, C. Deep perturbation learning: Enhancing the network performance via image perturbations. In Proceedings of the International Conference on Machine Learning, PMLR, Honolulu, HI, USA, 23–29 July 2023; pp. 32273–32287.
40. Li, C.; Zhang, D.; Huang, W.; Zhang, J. Cross contrasting feature perturbation for domain generalization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 1327–1337.
41. Dreyer, M.; Achibat, R.; Wiegand, T.; Samek, W.; Lapuschkin, S. Revealing hidden context bias in segmentation and object detection through concept-specific explanations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 3828–3838.
42. Zhou, J.; He, Z.; Lam, K.M.; Wang, Y.; Zhang, W.; Guo, C.; Li, C. AMSP-UOD: When vortex convolution and stochastic perturbation meet underwater object detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 7659–7667.
43. Fan, Q.; Segu, M.; Tai, Y.W.; Yu, F.; Tang, C.K.; Schiele, B.; Dai, D. Towards robust object detection invariant to real-world domain shifts. In Proceedings of the Eleventh International Conference on Learning Representations (ICLR 2023), Kigali, Rwanda, 1–5 May 2023.
44. Zhou, K.; Yang, Y.; Qiao, Y.; Xiang, T. Mixstyle neural networks for domain generalization and adaptation. *Int. J. Comput. Vis.* **2024**, *132*, 822–836. [[CrossRef](#)]
45. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)] [[PubMed](#)]
46. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
47. 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](#)]
48. Li, J.; Qu, C.; Shao, J. Ship detection in SAR images based on an improved faster R-CNN. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA), Beijing, China, 13–14 November 2017; pp. 1–6.
49. Wang, Y.; Wang, C.; Zhang, H.; Dong, Y.; Wei, S. A SAR dataset of ship detection for deep learning under complex backgrounds. *Remote Sens.* **2019**, *11*, 765. [[CrossRef](#)]
50. Xian, S.; Zhirui, W.; Yuanrui, S.; Wenhui, D.; Yue, Z.; Kun, F. AIR-SARShip-1.0: High-resolution SAR ship detection dataset. *J. Radars* **2019**, *8*, 852–863.
51. 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; pp. 9759–9768.

52. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. CenterNet: Keypoint Triplets for Object Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.
53. Tian, Z.; Chu, X.; Wang, X.; Wei, X.; Shen, C. Fully convolutional one-stage 3d object detection on lidar range images. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 34899–34911.
54. 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. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21002–21012.
55. Ross, T.Y.; Dollár, G. Focal loss for dense object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2980–2988.
56. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
57. Lu, X.; Li, B.; Yue, Y.; Li, Q.; Yan, J. Grid r-cnn. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7363–7372.
58. Pang, J.; Chen, K.; Shi, J.; Feng, H.; Ouyang, W.; Lin, D. Libra R-CNN: Towards Balanced Learning for Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
59. Zhang, H.; Li, F.; Liu, S.; Zhang, L.; Su, H.; Zhu, J.; Ni, L.M.; Shum, H.Y. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv* **2022**, arXiv:2203.03605.
60. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. Dets beat yolos on real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 16965–16974.

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.