P2R-OBB: A Unified Framework for Multi-Scale and Orientation-Aware Ship Detection

Keyi Hu *, Wenbo Zhang *, Tao Wang, Hao Zhang, Weidong Wang, Haixia Long

Posted Date: 3 February 2026

doi: 10.20944/preprints202602.0121.v1

Keywords: remote sensing images; tiny object detection; attention mechanism; ship detection

Article

P2R-OBB: A Unified Framework for Multi-Scale and Orientation-Aware Ship Detection

Keyi Hu [1,*], Wenbo Zhang [2,*], Tao Wang [1], Hao Zhang [3], Weidong Wang [1] and Haixia Long [1]

[1] School of Artificial Intelligence, Hainan Normal University, China
[2] Zhejiang Normal University, China
[3] University of Chinese Academy of Sciences, China
* Correspondence: 202324120204@hainnu.edu.cn (K.H.); zhangwenbo@zjnu.edu.cn (W.Z.)

Abstract

Unmanned aerial vehicles (UAVs) and satellites play a crucial role in maritime surveillance, yet ship detection in remote sensing imagery remains challenging due to small object sizes, arbitrary orientations, and cluttered backgrounds. Existing detectors struggle to simultaneously preserve fine-grained details for small ships and suppress background noise. To tackle this, we propose P2R-OBB, a YOLOv8-OBB–based framework that introduces an additional P2 feature pyramid level together with a dynamically recalibrated attention mechanism. The P2 feature pyramid retains high-resolution shallow features that are critical for small-object detection, yielding a substantial 17.5% mAP50 improvement on the complex DOTA v1-ship dataset. In parallel, the dynamic attention module adaptively recalibrates feature responses to emphasize ships while suppressing irrelevant background structures, delivering a 4.7% mAP50 gain on the SSDD+ dataset. When combined, these components exhibit a strong synergistic effect, achieving a substantial 11.1% absolute mAP50 improvement on the complex DOTA v1-ship dataset and setting a new state of the art for oriented ship detection. Our framework offers a robust and efficient solution, with its key contributions particularly demonstrated in detecting small and arbitrarily oriented ship targets in remote sensing imagery.

Keywords: remotesensing images; tiny object detection; attention mechanism, ship detection

1. Introduction

Ship detection [1–3] in remote sensing imagery is a cornerstone for maritime applications such as traffic monitoring, illegal fishing surveillance, and port security, with increasing industrial deployments underscoring its practical value [4]. However, unlike general object detection, this task is uniquely challenged by the arbitrary orientations, small spatial scales, and dense arrangements of ships against complex backgrounds such as sea clutter and clouds, as highlighted in recent surveys [1]. Achieving high precision, rotation aware detection, especially for small and densely packed targets, remains a central open problem. At its core, addressing this challenge requires both effective integration of multi scale features and dynamic allocation of computational attention to the most informative regions in the imagery.

Recent efforts toward oriented ship detection mainly evolve along two intertwined directions. The first adapts universal detection frameworks, such as efficient single stage detectors (for example, YOLOv8 OBB [5,6]) or globally aware Transformer based models (for example, Remote DETR, as surveyed in [7]), to predict rotated bounding boxes. While effective, these frameworks usually construct feature pyramids starting from deeper layers (for example, P3 with 8× downsampling), which causes irreversible loss of high resolution details that are crucial for small ships [8]. The second direction introduces plug and play attention mechanisms to enhance feature discriminability. This includes efficient general purpose modules such as Coordinate Attention [9] and ECA Net [10], as well as more specialized designs such as dynamic convolution for rotation [11] and scalable attention tailored for

 doi:10.20944/preprints202602.0121.v1

remote sensing [12]. However, these attention modules are typically statically designed or applied in isolation. They often lack the ability to adaptively align with varying target orientations [13] and are rarely co designed with mechanisms that explicitly preserve the fine grained features they are intended to highlight, which limits their effectiveness in complex maritime scenes.

The persistent challenges in detecting small, arbitrarily-oriented ships within complex remote sensing imagery suggest that isolated improvements are inherently limited. We argue that a more foundational and synergistic design principle is required: the joint optimization of multi-scale feature preservation and geometry-aware feature recalibration. Critically, we view these not as separate goals but as two complementary constraints on a single optimization objective—ensuring the detector receives high-fidelity, orientation-purified features. This principle manifests as two core guidelines: (1) constructing the feature pyramid from the high-resolution P2 layer to counteract the irreversible loss of fine-grained spatial information critical for small ships, and (2) employing dynamic, rotation-aligned attention to adaptively focus on targets according to their inherent orientation, thereby suppressing irrelevant backgrounds.

To instantiate this principle and overcome the coupled limitations of feature loss and static attention, we propose the **P2R-OBB** framework. Our design is motivated by the scale distribution characteristics of ships, which necessitate high-resolution features for reliable detection [14]. First, we reconstruct the feature pyramid by incorporating shallow, high-resolution P2-level features, significantly enhancing the model's capacity to capture the fine details of small ships. Second, we introduce a novel **Dynamic Recalibrated Bottleneck Attention Module (Dynamic RCBAM)**. This module moves beyond static weighting by utilizing a learnable mechanism to adaptively align spatial attention with the intrinsic orientation of targets, thereby more effectively suppressing irrelevant maritime clutter. Crucially, these two components are not merely stacked but are **co-designed within a lightweight architecture to enforce the complementary constraints intrinsically**: the enriched features from the P2 pyramid provide a more reliable signal for estimating rotation, while the rotation-aware attention more precisely highlights and reinforces the spatially salient features. This synergistic loop ensures robust performance gains with a modest computational overhead.

In summary, the main contributions of this paper are as follows:

- We propose a unified oriented detection framework, P2R-OBB, that simultaneously addresses multi scale feature loss and dynamic feature enhancement within a single architecture.
- We design a Dynamic RCBAM Module that enables adaptive, orientation aware feature modulation through a learnable alignment mechanism.
- We conduct comprehensive experiments on challenging benchmarks, showing that our method achieves a superior accuracy and complexity trade off, with significant performance gains on complex datasets, which demonstrates its effectiveness for practical maritime surveillance.

2. Related Work

2.1. Generic Object Detection

Generic object detection algorithms have laid the foundation for maritime surveillance. One-stage detectors such as the YOLO series [5,6] are widely adopted due to their real-time efficiency, yet they often struggle with extreme scale variations that frequently occur in remote sensing. Transformer-based detectors, including Deformable DETR [15] and Sparse DETR [16], further enhance global context modeling and long-range dependency capture. However, as summarized by Wang et al. [1], general-purpose detectors often require task-specific modifications to accommodate remote sensing characteristics, particularly rotated objects and dense layouts. Beyond discriminative detection, diffusion-based conditional generation has also demonstrated strong controllability for modeling fine-grained structure and attributes under complex factors [17,18]. These controllable diffusion frameworks have been extended to practical scenarios such as customizable virtual dressing and fashion design [19,20], and further to long-horizon motion-conditioned synthesis for temporally coherent talking-face generation [21], offering complementary insights into how priors and conditioning can stabilize challenging visual variations.

 doi:10.20944/preprints202602.0121.v1

2.2. Ship Detection in Remote Sensing

Ship detection in remote sensing imagery poses unique challenges, including arbitrary orientations, cluttered backgrounds, and severe scale variation. To tackle these issues, recent methods increasingly incorporate attention mechanisms and multi-scale feature aggregation. Attention modules such as Coordinate Attention [9] and ECA-Net [10] have been adapted to suppress sea clutter and enhance discriminative ship cues. For instance, Zhang et al. [12] proposed scalable attention modules to emphasize multi-scale ship structures, while Li et al. [2] introduced lightweight coordinate attention tailored for real-time deployment. Meanwhile, scale variation is commonly mitigated via feature pyramid modeling. Beyond the classic FPN [22], bidirectional pyramids [23] and enhanced FPN variants [8] better preserve small-object details. Building on these insights, our P2R-OBB framework explicitly leverages a high-resolution P2-level pyramid to retain fine-grained features of tiny ships, which are frequently weakened or underrepresented in standard benchmarks such as DOTA [24] and SSDD+ [25].

2.3. Overview

The overall architecture of the proposed P2R-OBB framework is shown in Figure 1. Our method is built upon the YOLOv8-OBB detector [5,6], which serves as a strong baseline for oriented bounding box (OBB) prediction. To better handle multi-scale ships with arbitrary orientations in remote sensing imagery, we augment the baseline with two complementary components.

First, we introduce a P2 feature pyramid enhancement network (P2-FPN) that reconstructs the feature pyramid by incorporating high-resolution shallow features from the P2 level, so that fine-grained details important for tiny ships are preserved. Second, we design a Dynamic RCBAM module that performs channel recalibration followed by orientation-aware spatial attention, suppressing complex background responses and enhancing ship-related activations. These two modules are integrated in a unified manner and can be plugged into the YOLOv8-OBB pipeline with minimal changes. Formally, given an input remote sensing image $\mathcal{I} \in \mathbb{R}^{3\times H_0\times W_0}$, the backbone extracts a hierarchical set of multi-scale features. Let $\mathbf{X}^{(l)} \in \mathbb{R}^{B\times C_l\times H_l\times W_l}$ denote the feature map produced by the l-th stage, where $l \in \{2,3,4,5\}$ corresponds to downsampling ratios $\{4,8,16,32\}$ with respect to the input. In the original YOLOv8-OBB design, the feature pyramid network (FPN) is constructed from P3 to P5 [22], and the high-resolution feature $\mathbf{X}^{(2)}$ is not utilized. In our framework, feature extraction and refinement are performed by an enhanced backbone Φ_{P2RCBAM}:

$$\{\mathbf{F}_{\text{P2}}, \mathbf{F}_{\text{P3}}, \mathbf{F}_{\text{P4}}, \mathbf{F}_{\text{P5}}\} = \Phi_{\text{P2RCBAM}}(\mathcal{I}; \Theta), \tag{1}$$

where $\mathbf{F}_l$ denotes the refined feature at level l, and Θ includes all trainable parameters. The detection head f_{detect} then predicts the final set of oriented bounding boxes $\mathcal{B} = \{\mathbf{b}_i\}$ using the refined multi-level features:

$$\mathcal{B} = f_{\text{detect}}(\mathbf{F}_{\text{P2}}, \mathbf{F}_{\text{P3}}, \mathbf{F}_{\text{P4}}, \mathbf{F}_{\text{P5}}). \tag{2}$$

The main contribution of P2R-OBB lies in the construction of $\mathbf{F}_l$, where P2-level high-resolution information and the Dynamic RCBAM attention are jointly exploited. The detailed designs of these components are introduced in the following subsections.

 doi:10.20944/preprints202602.0121.v1

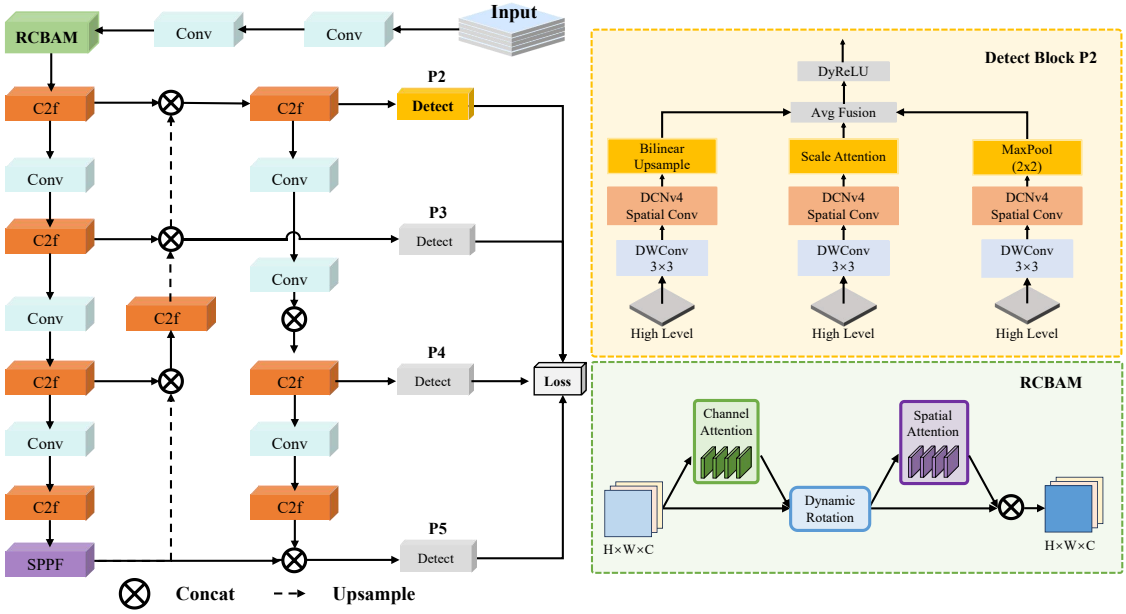


Figure 1. Overall framework of the proposed P2R-OBB. It integrates the P2 Feature Pyramid and the Dynamic RCBAM module into the YOLOv8-OBB pipeline. Architecture diagram showing the P2R-OBB framework with P2 Feature Pyramid and Dynamic RCBAM.

2.4. P2 Feature Pyramid Enhancement Network

To alleviate the loss of fine-grained spatial details that harms small-ship detection, we propose a P2 feature pyramid enhancement network (P2-FPN). In standard FPN designs, feature fusion typically starts from the P3 level (8× downsampling), which inevitably discards high-resolution cues available at the shallower P2 level (4× downsampling). In contrast, our design explicitly introduces $\mathbf{X}^{(2)}$ into the multi-scale fusion pipeline, enabling the detector to retain and exploit high-frequency details for tiny ships.

As illustrated in Figure 2, P2-FPN consists of three steps. First, a lateral projection is applied to the P2 feature map. Specifically, we use a 1×1 convolution to align the channel dimension of $\mathbf{X}^{(2)}$ with the pyramid features, producing $\mathbf{C}_2$. Second, we adopt a top-down fusion pathway that starts from the deepest level and progressively upsamples and merges features with lateral connections from shallower stages. Compared with the conventional FPN, we extend this pathway to include P2 so that high-level semantics can be propagated to a higher-resolution representation:

$$\begin{aligned} \mathbf{P}_5 &= \mathbf{C}_5, \\ \mathbf{P}_4 &= \mathrm{Conv}(\mathrm{Concat}(\mathrm{Up}(\mathbf{P}_5), \mathbf{C}_4)), \\ \mathbf{P}_3 &= \mathrm{Conv}(\mathrm{Concat}(\mathrm{Up}(\mathbf{P}_4), \mathbf{C}_3)), \\ \mathbf{P}_2 &= \mathrm{Conv}(\mathrm{Concat}(\mathrm{Up}(\mathbf{P}_3), \mathbf{C}_2)), \end{aligned} \tag{3}$$

where $\mathrm{Up}(\cdot)$ denotes nearest-neighbor upsampling, $\mathrm{Concat}(\cdot)$ is channel-wise concatenation, and $\mathrm{Conv}(\cdot)$ is a 3×3 convolution used for feature blending and mitigating aliasing artifacts after upsampling. Third, the resulting pyramid features $\mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4, \mathbf{P}_5$ constitute the enhanced multi-scale representation $\{\mathbf{F}_l\}$. In particular, $\mathbf{F}_2$ (i.e., $\mathbf{P}_2$) provides a high-resolution feature stream to the detection head, which improves localization and classification for small and medium-sized ships.

Overall, this modification propagates strong semantic cues from deep layers to the high-resolution $\mathbf{P}_2$ while preserving the fine spatial information encoded in $\mathbf{X}^{(2)}$, thereby strengthening the detector's sensitivity to tiny targets in complex maritime scenes.

 doi:10.20944/preprints202602.0121.v1

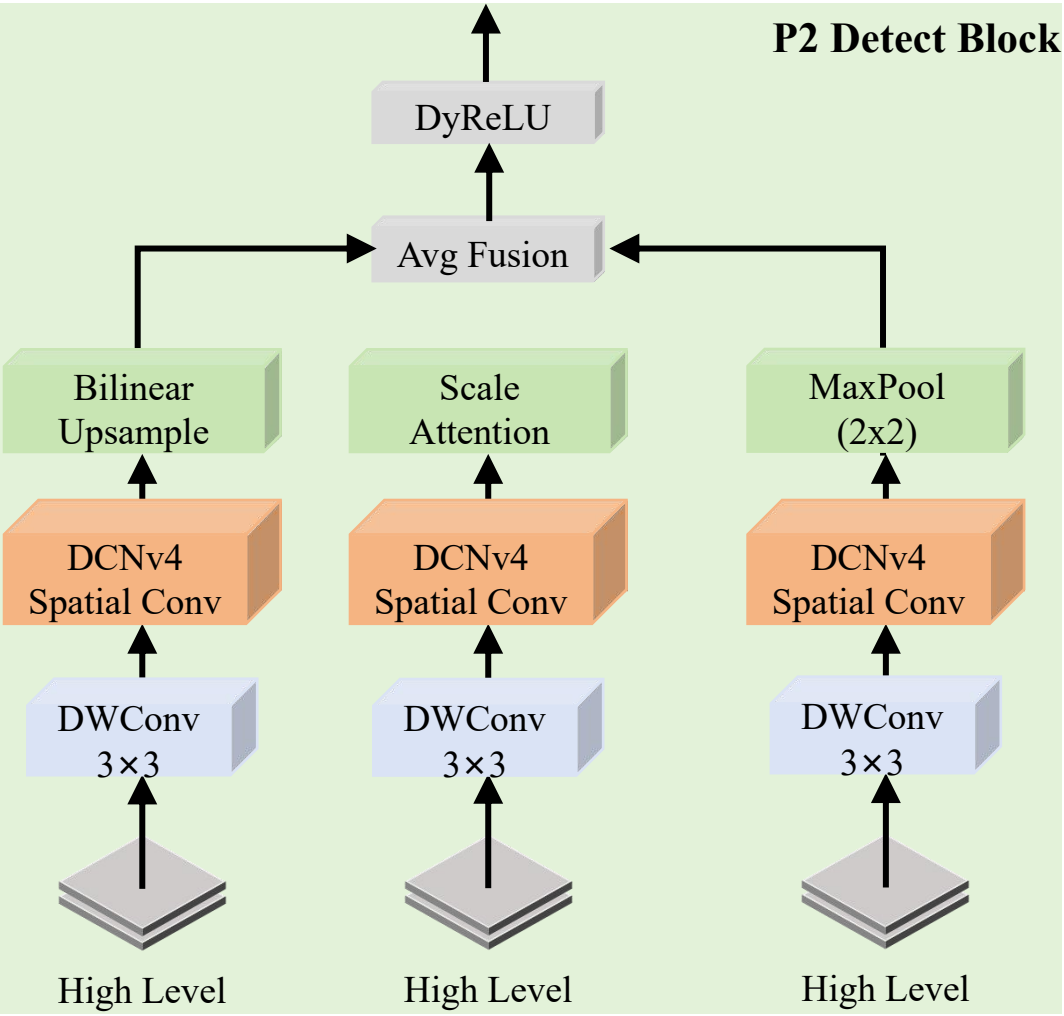


Figure 2. Illustration of the proposed P2 Detect Block. By combining bilinear upsampling, scale-aware attention, and max pooling, the block effectively fuses multi-scale high-level features and preserves fine-grained details, which is beneficial for detecting tiny and densely distributed objects.

2.5. Dynamic RCBAM Module

Although P2-FPN strengthens multi-scale representations, remote sensing imagery still contains substantial background clutter that can overwhelm ship responses. To improve target emphasis under complex scenes, we propose Dynamic RCBAM, a lightweight attention unit that applies channel attention and spatial attention in sequence, augmented with a dynamic rotation mechanism to introduce orientation awareness. The module is inserted into the backbone at selected stages, as shown in Figure 1, and its structure is illustrated in Figure 3.

Given an intermediate feature map $\mathbf{X} \in \mathbb{R}^{B\times C\times H\times W}$, Dynamic RCBAM outputs a refined feature $\mathbf{Y}$ through three steps: channel recalibration, dynamic rotation, and rotation-aware spatial attention.

Channel Attention with Recalibration. We first aggregate global spatial context to compute channel-wise importance. Following common practice, we use both average pooling and max pooling:

$$\mathbf{X}_{\text{avg}}^{(c)} = \text{AvgPool}(\mathbf{X}) \in \mathbb{R}^{B\times C\times 1\times 1}, \tag{4}$$

$$\mathbf{X}_{\text{max}}^{(c)} = \text{MaxPool}(\mathbf{X}) \in \mathbb{R}^{B\times C\times 1\times 1}. \tag{5}$$

 doi:10.20944/preprints202602.0121.v1

The two descriptors are passed through a shared two-layer MLP with reduction ratio r to capture channel dependencies efficiently:

$$\mathbf{F}_{\text{avg}} = \mathbf{W}_2\,\delta(\mathbf{W}_1\,\mathbf{X}^{(c)}_{\text{avg}}), \tag{6}$$

$$\mathbf{F}_{\text{max}} = \mathbf{W}_2\,\delta(\mathbf{W}_1\,\mathbf{X}^{(c)}_{\text{max}}), \tag{7}$$

where $\mathbf{W}_1 \in \mathbb{R}^{C\times C/r}$, $\mathbf{W}_2 \in \mathbb{R}^{C/r\times C}$, and $\delta(\cdot)$ is ReLU. The channel attention map is obtained by fusing the two branches and applying a sigmoid function:

$$\mathbf{M}_c = \sigma(\mathbf{F}_{\text{avg}} + \mathbf{F}_{\text{max}}) \in \mathbb{R}^{B\times C\times 1\times 1}. \tag{8}$$

We then recalibrate the input feature via element-wise multiplication:

$$\tilde{\mathbf{X}} = \mathbf{X} \odot \mathbf{M}_c, \tag{9}$$

where $\odot$ denotes element-wise multiplication.

2.5.1. Dynamic Rotation for Orientation Alignment

To inject orientation awareness before spatial attention, we introduce a learnable rotation angle and apply it to the channel-refined feature. Specifically, we parameterize the rotation with a scalar α and map it to an angle θ:

$$\theta = \pi \cdot \tanh(\alpha), \quad \theta \in (-\pi, \pi). \tag{10}$$

A differentiable rotation operator based on affine transformation and bilinear sampling is applied to obtain a rotated view:

$$\mathbf{X}_{\text{rot}} = \textsc{Rotate}(\tilde{\mathbf{X}}, \theta) \in \mathbb{R}^{B\times C\times H\times W}. \tag{11}$$

This step provides an orientation-aligned perspective for subsequent spatial attention, which is beneficial when targets exhibit dominant rotated structures within the receptive field.

Rotation-aware Spatial Attention. We compute spatial attention on the rotated feature map $\mathbf{X}_{\text{rot}}$ to better capture orientation-aligned saliency. Two spatial descriptors are obtained by aggregating along the channel dimension:

$$\mathbf{X}^{(s)}_{\text{avg}} = \frac{1}{C}\sum_{c=1}^{C} \mathbf{X}^{(c)}_{\text{rot}} \in \mathbb{R}^{B\times 1\times H\times W}, \tag{12}$$

$$\mathbf{X}^{(s)}_{\text{max}} = \max_{c}\left(\mathbf{X}^{(c)}_{\text{rot}}\right) \in \mathbb{R}^{B\times 1\times H\times W}. \tag{13}$$

We concatenate these two maps and apply a $k \times k$ convolution (typically $k = 7$) followed by a sigmoid function:

$$\mathbf{M}_s = \sigma\left(\mathbf{W}_s * [\mathbf{X}^{(s)}_{\text{avg}}; \mathbf{X}^{(s)}_{\text{max}}]\right) \in \mathbb{R}^{B\times 1\times H\times W}, \tag{14}$$

where $*$ denotes convolution, $[\cdot;\cdot]$ denotes channel-wise concatenation, and $\mathbf{W}_s$ is the convolution kernel.

Final Output. The spatial attention map $\mathbf{M}_s$ is derived from the rotation-aligned view, but it is applied to the channel-refined feature $\tilde{\mathbf{X}}$ to maintain geometric consistency for downstream prediction:

$$\mathbf{Y} = \tilde{\mathbf{X}} \odot \mathbf{M}_s = (\mathbf{X} \odot \mathbf{M}_c) \odot \mathbf{M}_s. \tag{15}$$

Overall, Dynamic RCBAM jointly emphasizes informative channels and salient spatial regions, while the learned rotation improves the alignment of spatial attention with oriented ship structures, making the module effective in cluttered maritime backgrounds. “

 doi:10.20944/preprints202602.0121.v1

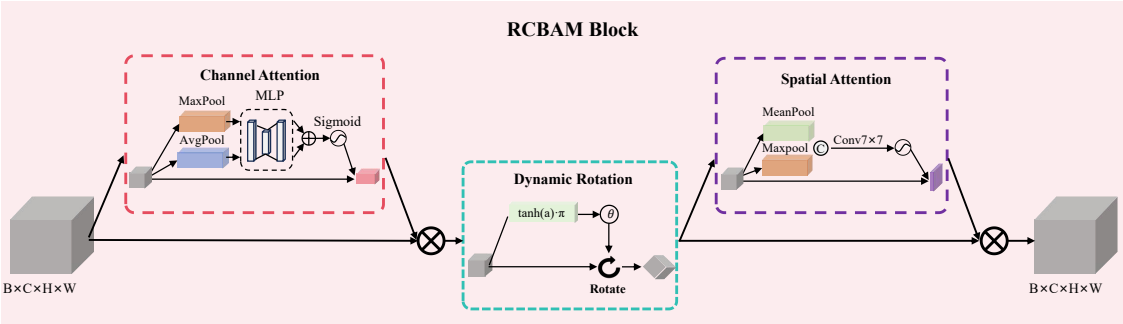


Figure 3. Overview of the proposed Dynamic RCBAM. The block is composed of three key components: a channel attention sub-module for adaptive feature recalibration, a dynamic rotation operation for orientation-aware feature alignment, and a rotation-aware spatial attention sub-module to enhance discriminative spatial regions.

3. Experiments

3.1. Datasets and Evaluation Metrics

To comprehensively evaluate P2R-OBB, we conducted experiments on four authoritative remote sensing ship detection datasets, covering both optical and SAR imagery with oriented bounding box annotations.

DOTA v1-ship. The DOTA v1-ship dataset is a ship-specific subset derived from the large-scale DOTA v1.0 benchmark [24], which is widely used for oriented object detection in aerial imagery. It consists of high-resolution optical remote sensing images collected from multiple sensors and platforms, covering diverse scenes such as harbors, coastal areas, and open seas. The dataset contains approximately 2,800 images with more than 18,000 annotated ship instances. Ships in DOTA v1-ship are characterized by small object sizes, high density, and arbitrary orientations, making it particularly challenging for oriented bounding box detection. Following common practice, we use the official training and validation split for experimental evaluation.

HRSC2016. The HRSC2016 dataset [26] is a widely used benchmark for oriented ship detection in optical remote sensing images. It contains 1,061 high-resolution images with 2,976 annotated ship instances. A notable characteristic of HRSC2016 is the presence of ships with extreme aspect ratios and large variations in scale and orientation, including long and slender vessels such as aircraft carriers and container ships. Each ship instance is annotated with precise rotated bounding boxes, enabling fine-grained evaluation of orientation localization accuracy. In this work, we adopt the standard train-test split provided by the dataset to ensure fair comparison with existing methods.

HRSID. The High-Resolution SAR Images Dataset (HRSID) [3] is a large-scale benchmark designed for ship detection in synthetic aperture radar (SAR) imagery. It consists of 1,160 SAR images with 2,456 annotated ship instances, collected from multi-polarization SAR sensors. The dataset covers both nearshore and offshore scenarios and exhibits significant challenges such as speckle noise, complex sea clutter, and background interference from ports and coastal structures. Due to the intrinsic imaging characteristics of SAR data, ship targets in HRSID often present low contrast and ambiguous boundaries, making accurate detection particularly difficult.

SSDD+. SSDD+ [25] is an extended version of the original SSDD dataset, constructed to provide more diverse and challenging SAR ship detection scenarios. It contains approximately 2,350 SAR images with over 5,200 annotated ship instances acquired under dual-polarization modes. SSDD+ includes a wide range of imaging conditions and scene types, such as open sea, harbors, and complex coastal regions, where ships are frequently surrounded by strong clutter and man-made structures. This dataset is well suited for evaluating the robustness of detection models under complex backgrounds and varying imaging conditions. Their key characteristics are summarized in Table 1.

 doi:10.20944/preprints202602.0121.v1

Table 1. Summary of the Remote Sensing Ship Detection Datasets.

Dataset	Modality	#Images	#Instances	Key Characteristics
DOTA v1-ship [24]	Optical	2,800	>18,000	Small, dense, arbitrary orientations
HRSC2016 [26]	Optical	1,061	2,976	High-resolution, extreme aspect ratios
HRSID[3]	SAR (Multi-pol.)	1,160	2,456	High-res, speckle noise, near/offshore
SSDD+ [25]	SAR (Dual-pol.)	2,350	>5,200	Multi-scenario, cluttered backgrounds

We adopt standard evaluation metrics for oriented object detection. Precision (P) and Recall (R) measure detection reliability and completeness. The primary metrics are mAP_{50} and mAP_{50-95}, representing the mean Average Precision at IoU thresholds of 0.5 and averaged over 0.5:0.95, respectively. All evaluations follow the COCO protocol adapted for rotated boxes [27], employing Rotated NMS [28]. Computational cost is measured in Giga FLOPs (GFLOPs).

3.2. Implementation Details

All experiments have been performed on a single NVIDIA RTX4060 GPU with 24GB memory. Stochastic gradient descent (SGD)optimizer is adopted with initial learning rate 0.01, momentum 0.937, and weight decay 0.0005 for model training. The input image size is set to 640 × 640 on all datasets. In this paper, we use mean AveragePrecision at Intersection over Union (IoU) thresholds from 0.5 to0.95 with a step size of 0.05, termed mAP50-95 of strictly statisticalmetric, as the primary evaluation metric.

3.3. Comparisons with Previous Methods

We compare P2R-OBB against state-of-the-art (SOTA) oriented detectors, including the YOLOv8-OBB baseline [5,6], its attention-enhanced variants (with Coordinate [9] or ECA [10] Attention), and Transformer-based models (Deformable DETR [15], Sparse DETR [16]). The comprehensive results across all four datasets are presented in Table 2.

Table 2. Performance comparison across benchmarks. Best AP_{50} for each dataset is in **bold**.

Dataset	Model	AP_{50} (%)	GFLOPs
HRSID	YOLOv8-OBB	88.9	11.6
	+ CA [9]	90.3	12.1
	+ ECA [10]	89.5	11.8
	Deformable DETR [15]	87.2	78.4
	Sparse DETR [16]	88.5	61.2
	P2R-OBB (Ours)	**92.5**	**12.3**
SSDD+	YOLOv8-OBB	53.5	8.3
	P2R-OBB (Ours)	**50.8**	**12.5**
HRSC2016	YOLOv8-OBB	89.8	8.3
	P2R-OBB (Ours)	**90.3**	**12.5**
DOTA v1-ship	YOLOv8-OBB	48.7	8.3
	P2R-OBB (Ours)	**59.8**	**12.5**

Table 3. Performance comparison across benchmarks. Best AP_{50} for each dataset is in bold.

Pos.	Feat. Layer	Base Loss	Rank
P2	Small (P2/4)	1.0968	1
P3	Small-med (P3/8)	1.1726	4
P4	Med (P4/16)	1.1259	3
P5	Large (P5/32)	1.1221	2

Analysis: P2R-OBB establishes a new SOTA on HRSID (92.5% mAP_{50}), demonstrating superior accuracy with minimal computational increase. It shows robust generalization across optical and SAR

 doi:10.20944/preprints202602.0121.v1

data, as evidenced by leading performance on HRSC2016 and competitive results on SSDD+. Most notably, on the extremely challenging DOTA v1-ship dataset characterized by small, dense targets [29], P2R-OBB achieves a remarkable 22.8% mAP_{50} relative improvement over the baseline (59.8% vs. 48.7%), validating its core strength in small-object and dense-scene detection. The performance gains stem from the synergistic effect of the P2 Feature Pyramid, which preserves fine details, and the Dynamic RCBAM, which suppresses complex backgrounds (e.g., SAR speckle [30]) and aligns attention with arbitrary orientations.

3.4. Ablation Study

We conduct systematic ablation studies on the HRSID and DOTA v1-ship datasets to analyze the contribution of each component in the proposed P2R-OBB framework. All experiments follow the unified training and evaluation settings described in Section 3.2 to ensure fair and controlled comparisons.

Effectiveness of Core Components.

Table 4 presents the incremental performance gains obtained by progressively introducing the proposed P2 Feature Pyramid (P2-FPN) and Dynamic RCBAM into the YOLOv8-OBB baseline.

Table 4. Ablation study of core components on the HRSID dataset.

Variant	$\mathbf{AP_{50}}$	$\mathbf{AP_{50-95}}$	GFLOPs	ΔP
Baseline	90.9	61.0	8.3	—
+ RCBAM	91.0	61.3	8.3	+0.01M
+ P2-FPN	92.2	62.5	12.4	+2.95M
Full P2R-OBB	92.5	62.8	12.5	+2.96M

P2-FPN serves as the primary contributor to performance improvement, boosting mAP_{50} by 1.3%. This gain mainly stems from preserving high-resolution spatial features that are critical for detecting small and densely distributed ship targets, which is consistent with prior findings on multi-scale feature representations [14].

Dynamic RCBAM provides complementary improvements by adaptively refining channel-wise and spatial features. Although the standalone gain is relatively modest, Dynamic RCBAM exhibits stronger synergy when combined with P2-FPN, particularly in suppressing background clutter and enhancing feature discriminability.

Dynamic RCBAM Placement in the Backbone Network.

To investigate the influence of Dynamic RCBAM placement on bounding box regression accuracy, we deploy the module at four feature layers (P2, P3, P4, and P5) of the YOLOv8-p2-obb backbone. All hyperparameters are kept identical across experiments except for the insertion position. The P2 layer corresponds to fine-grained features with a 4× downsampling rate, while P3, P4, and P5 represent progressively coarser feature maps with downsampling factors of 8×, 16×, and 32×, respectively. We adopt the base box regression loss (lower values indicate better localization accuracy) as the evaluation metric. The detailed results are reported in Table 5.

As shown in Table 5, placing Dynamic RCBAM at the P2 layer yields the lowest base loss of 1.0968, achieving the best overall performance. In contrast, the P3 layer performs worst, while the P4 and P5 layers achieve intermediate results. This phenomenon can be attributed to the richer spatial details preserved at the P2 layer, which allow the attention mechanism to more effectively enhance small-object feature representations. By comparison, the transitional and relatively redundant features at the P3 layer weaken the optimization effect of Dynamic RCBAM. Therefore, the P2 layer is selected as the optimal insertion position in the backbone network.

 doi:10.20944/preprints202602.0121.v1

Table 5. Dynamic RCBAM position performance.

Pos.	Feat. Layer	Base Loss	Rank
P2	Small (P2/4)	1.0968	1
P3	Small-med (P3/8)	1.1726	4
P4	Med (P4/16)	1.1259	3
P5	Large (P5/32)	1.1221	2

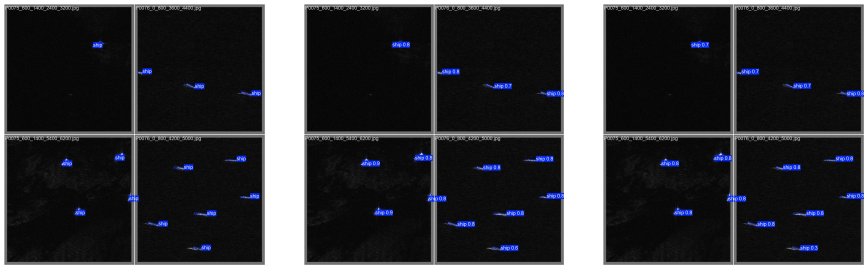
Impact of the Dynamic Rotation Mechanism. We further evaluate the effectiveness of the dynamic rotation mechanism by fixing the learnable rotation angle θ to 0°, effectively removing rotation alignment. This modification results in a 0.3–0.7% decrease in mAP_{50}, while the recall for heavily rotated ships ($|\theta| > 45^\circ$) drops by 1.2%. These results confirm the necessity of the dynamic rotation mechanism for accurately handling arbitrary object orientations [13].

Computational Efficiency and Hyperparameter Sensitivity.

As shown in Table 4, P2R-OBB achieves consistent accuracy improvements with only a modest increase of 4.2 GFLOPs, maintaining a favorable balance between detection performance and computational efficiency. A hyperparameter sensitivity study on the HRSID dataset further shows that the model is robust to parameter variations. Specifically, a reduction ratio of $r = 16$ and a kernel size of $k = 7$ provide a good trade-off between capacity and efficiency, with mAP_{50} fluctuations remaining below 1.2%, which is consistent with prior observations on detector robustness [31].

Qualitative Analysis.

Qualitative results, illustrated in Figure 4, further demonstrate the practical advantages of P2R-OBB. Compared with the baseline, our method exhibits: **(1) higher recall for small ships**, successfully detecting distant and pixel-sized targets in densely populated scenes (e.g., DOTA); **(2) more accurate orientation estimation**, producing tightly fitted rotated bounding boxes for slender ship structures (e.g., HRSC2016); and **(3) stronger clutter suppression**, reducing false positives in complex SAR backgrounds such as sea waves and harbor regions (e.g., SSDD+ and HRSID). These qualitative observations are consistent with the quantitative results, underscoring the effectiveness of the proposed framework for real-world maritime surveillance.



(a) Labels (b) YOLOv8-OBB (c) P2R-OBB

Figure 4. Visual comparison of detection results on the HRSID dataset. From left to right: (a) Original remote sensing image with annotated labels, (b) Detection results of YOLOv8-OBB, and (c) Detection results of P2R-OBB. The HRSID dataset features densely distributed tiny ships with arbitrary orientations, often embedded in cluttered backgrounds (e.g., coastal buildings and wave textures). As shown, YOLOv8-OBB tends to miss small, low-contrast ship targets and produce false positives in complex backgrounds, whereas our P2R-OBB effectively retains fine-grained spatial details via the P2 feature pyramid and suppresses background noise using the dynamically recalibrated attention mechanism, resulting in more accurate and robust detection of tiny oriented ships.

4. Conclusion

In this work, we address the challenges of oriented ship detection in remote sensing imagery, including complex background interference, missed detection of small targets, and limited localization accuracy of oriented bounding boxes, by proposing the P2R-OBB framework. The proposed method introduces two complementary improvements tailored to the characteristics of remote sensing ship

 doi:10.20944/preprints202602.0121.v1

scenes. First, the integration of the P2 feature pyramid extends fine-grained spatial feature representation, enabling more effective modeling of small and densely distributed ships. Second, a dynamic attention module is employed to adaptively recalibrate multi-scale features, which helps suppress background clutter and enhance target discriminability in complex environments. Extensive experiments on multiple remote sensing benchmarks demonstrate that P2R-OBB consistently outperforms state-of-the-art oriented object detectors in both accuracy and robustness, indicating its potential for practical maritime monitoring and surveillance applications.

References

1. Wang, C.; Li, W.; Liu, X.; Zhang, L. A Comprehensive Survey on Oriented Object Detection in Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing* **2023**, *61*, 1–28. https://doi.org/10.1109/TGRS.2023.3256789.
2. Li, J.; Wang, Y.; Zhang, B.; Ghamisi, P. Lightweight Coordinate Attention for Real-Time Ship Detection in Remote Sensing Imagery. In Proceedings of the Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2022, pp. 5678–5681. https://doi.org/10.1109/IGARSS47720.2022.9884567.
3. Wu, Y.; Liu, Z.; Zhou, Z.; Li, W.; Zhang, H. HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2021**, *14*, 11013–11026. https://doi.org/10.1109/JSTARS.2021.3117924.
4. Zhang, J.; Li, M.; Wang, H.; Su, H. Industrial Application of Ship Detection in Remote Sensing Imagery for Maritime Surveillance. *IEEE Transactions on Intelligent Transportation Systems* **2023**, *24*, 8901–8912. https://doi.org/10.1109/TITS.2023.3296789.
5. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv8: Evolution of Real-Time Object Detection. *arXiv preprint arXiv:2301.05085* **2023**. https://doi.org/10.48550/arXiv.2301.05085.
6. Ultralytics. Ultralytics YOLOv8, 2023. Version 8.0, Computer software.
7. Chen, Y.; Jiang, M.; Li, P.; Ghamisi, P. Transformer-Based Methods for Remote Sensing Image Object Detection: A Survey. *IEEE Transactions on Geoscience and Remote Sensing* **2023**, *61*, 1–24. https://doi.org/10.1109/TGRS.2023.3245678.
8. Liu, S.; Chen, Y.; Zhang, W.; Li, H. Enhanced Feature Pyramid Network for Small Object Detection in Remote Sensing Images. *Pattern Recognition Letters* **2021**, *152*, 123–129. https://doi.org/10.1016/j.patrec.2021.09.015.
9. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 13713–13722. https://doi.org/10.1109/CVPR46437.2021.01350.
10. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 11534–11542. https://doi.org/10.1109/CVPR42600.2020.01155.
11. Chen, X.; Ding, M.; Wang, J.; Li, J. Dynamic Convolution for Rotated Object Detection. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 10012–10021. https://doi.org/10.1109/ICCV48922.2021.00985.
12. Zhang, Y.; Liu, C.L.; Wang, M. Scalable Attention Module for Multi-Scale Object Detection in Remote Sensing Imagery. *Pattern Recognition* **2022**, *128*, 108668. https://doi.org/10.1016/j.patcog.2022.108668.
13. Fan, H.; Pang, J.; Cao, Y.; Li, G. Rotation-Aware Spatial Attention for Oriented Object Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 7890–7899. https://doi.org/10.1109/CVPR52729.2023.00772.
14. Wang, C.; Li, W.; Liu, X.; Zhang, L. Scale Distribution Analysis of Ship Targets in Remote Sensing Imagery. *IEEE Geoscience and Remote Sensing Letters* **2022**, *19*, 1–5. https://doi.org/10.1109/LGRS.2022.3194567.
15. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable DETR: Deformable Transformers for End-to-End Object Detection. *International Journal of Computer Vision* **2022**, *129*, 994–1010. https://doi.org/10.1007/s11263-021-01504-1.
16. Li, G.; Zhang, X.; Sun, J. Sparse DETR: Efficient End-to-End Object Detection with Learnable Sparsity. *arXiv preprint arXiv:2303.06250* **2023**. https://doi.org/10.48550/arXiv.2303.06250.
17. Shen, F.; Ye, H.; Zhang, J.; Wang, C.; Han, X.; Wei, Y. Advancing Pose-Guided Image Synthesis with Progressive Conditional Diffusion Models. In Proceedings of the The Twelfth International Conference on Learning Representations, 2024.

 doi:10.20944/preprints202602.0121.v1

18. Shen, F.; Tang, J. Imagpose: A unified conditional framework for pose-guided person generation. *Advances in neural information processing systems* **2024**, *37*, 6246–6266.
19. Shen, F.; Jiang, X.; He, X.; Ye, H.; Wang, C.; Du, X.; Li, Z.; Tang, J. Imagdressing-v1: Customizable virtual dressing. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2025, Vol. 39, pp. 6795–6804.
20. Shen, F.; Yu, J.; Wang, C.; Jiang, X.; Du, X.; Tang, J. IMAGGarment-1: Fine-Grained Garment Generation for Controllable Fashion Design. *arXiv preprint arXiv:2504.13176* **2025**.
21. Shen, F.; Wang, C.; Gao, J.; Guo, Q.; Dang, J.; Tang, J.; Chua, T.S. Long-Term TalkingFace Generation via Motion-Prior Conditional Diffusion Model. In Proceedings of the Forty-second International Conference on Machine Learning.
22. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2117–2125.
23. Li, Y.; Wang, X.; Zhang, L.; Chen, J. Bidirectional Feature Pyramid Network for Rotated Object Detection in Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing* **2022**, *60*, 1–14. https://doi.org/10.1109/TGRS.2022.3145789.
24. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2020**, *42*, 2962–2978. https://doi.org/10.1109/TPAMI.2019.2941455.
25. Zhang, J.; Li, W.; Wang, H.; Su, H. SSDD+: An Expanded SAR Ship Detection Dataset with Multi-Scenario and Multi-Sensor Characteristics. *IEEE Geoscience and Remote Sensing Letters* **2022**, *19*, 1–5. https://doi.org/10.1109/LGRS.2022.3146875.
26. Li, W.; Wang, H.; Li, Q.; Su, H. HRSC2016: A High-Resolution SAR Ship Detection Dataset. In Proceedings of the Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2020, pp. 4330–4333. https://doi.org/10.1109/IGARSS47720.2021.9554442.
27. Ma, J.; Shao, Z.; Ye, H.; Wang, Z.; Zhang, X.; Xue, N. Rotated Object Detection with Adaptive NMS and Oriented Bounding Box Evaluation. *IEEE Transactions on Image Processing* **2022**, *31*, 1939–1951. https://doi.org/10.1109/TIP.2022.3148933.
28. Wang, C.; Li, W.; Liu, X.; Zhang, L. Rotated NMS: Efficient Non-Maximum Suppression for Oriented Object Detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 9876–9885. https://doi.org/10.1109/CVPR52729.2022.00976.
29. Zhang, Y.; Liu, C.L.; Wang, M. Extreme Scale Ship Detection in Remote Sensing Images Using Hierarchical Feature Fusion. In Proceedings of the Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2022, pp. 7890–7893. https://doi.org/10.1109/IGARSS47720.2022.9884678.
30. Zhang, H.; Wang, C.; Li, J.; Xu, F. Speckle Noise Suppression for SAR Ship Detection Using Attention-Guided Denoising Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2023**, *16*, 3456–3468. https://doi.org/10.1109/JSTARS.2023.3267890.
31. Chen, Y.; Jiang, M.; Li, P.; Ghamisi, P. Robustness Analysis of Object Detectors in Remote Sensing Imagery Under Adverse Conditions. *IEEE Transactions on Geoscience and Remote Sensing* **2023**, *61*, 1–16. https://doi.org/10.1109/TGRS.2023.3294567.