

Article

YOSDet: A YOLO-Based Oriented Ship Detector in SAR Imagery

Chushi Yu , Oh-Soon Shin  and Yoan Shin * 

School of Electronic Engineering, Soongsil University, Seoul 06978, Republic of Korea; osshin@ssu.ac.kr (O.-S.S.)

* Correspondence: yashin@ssu.ac.kr

Highlights

What are the main findings?

- YOSDet, a YOLO-based oriented ship detector, effectively handles arbitrarily oriented ships in SAR imagery, achieving high detection accuracy across SSDD+, HRSID, and SRSD-v1.0 benchmarks.
- The model integrates a dynamic aggregation module (DAM), an objective-guided detection head (OGDH), and a localization quality estimator (LQE), improving prediction consistency under noisy SAR imaging conditions.

What are the implications of the main findings?

- The results demonstrate robust generalization for both inshore and offshore scenarios, making the framework suitable for real-time maritime surveillance.
- This work provides a practical framework for oriented ship detection in complex SAR environments, highlighting the potential of deep learning in remote sensing applications.

Abstract

Synthetic aperture radar (SAR) serves as a prominent remote sensing (RS) technology, permitting continuous maritime surveillance regardless of weather or time. Although deep learning-based detectors have achieved promising results in SAR imagery, the majority of current algorithms rely on axis-aligned bounding boxes, which are insufficient for accurately representing arbitrarily oriented ships, especially under speckle noise, complex coastal clutter, and real-time deployment constraints. To address this limitation, we propose a YOLO-based oriented ship detector (YOSDet). Specifically, a dynamic aggregation module (DAM) is incorporated into the backbone to enhance feature representation against non-stationary backscattering. An objective-guided detection head (OGDH) is developed to decouple classification and localization, complemented by a localization quality estimator (LQE) to calibrate classification confidence by mitigating the impact of scattering center shifts. Comparative evaluations conducted on three public SAR ship detection benchmarks validate the effectiveness of YOSDet. The proposed model outperforms existing detectors, achieving *mAP* scores of 96.8%, 88.5%, and 67.3% on the SSDD+, HRSID, and SRSD-v1.0 datasets, respectively. Furthermore, the consistency of our approach in both nearshore and offshore environments is confirmed through rigorous quantitative and qualitative assessments.

Keywords: oriented ship detection; SAR imagery; deep learning; you only look once

Academic Editor: Massimiliano Pieraccini

Received: 16 January 2026

Revised: 12 February 2026

Accepted: 17 February 2026

Published: 19 February 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.

1. Introduction

Remote sensing (RS) facilitates the acquisition of data regarding remote targets or environmental processes by recording backscattered electromagnetic energy via specialized

sensors [1,2]. Within the diverse landscape of RS technologies, synthetic aperture radar (SAR) stands out as an active imaging modality, offering reliable performance regardless of time or atmospheric conditions. Owing to its robustness against illumination variations and atmospheric interference, SAR is frequently utilized for ecological observation, oceanic monitoring, and military reconnaissance [3]. In particular, the precise identification of vessels within SAR data is indispensable for coastal security management, illegal fishing supervision, and maritime traffic management [4]. Despite these advantages, the coherent imaging mechanism of SAR inevitably introduces strong speckle noise and complex sea clutter, which obscure target boundaries and complicate precise localization and discrimination [5]. Moreover, ships in SAR scenes often exhibit arbitrary orientations, with varying scales and dense distributions, especially in ports and near-shore areas.

Recent advances in deep learning (DL) have fundamentally transformed object detection into end-to-end trainable frameworks capable of learning hierarchical representations directly from data [6]. Most state-of-the-art (SOTA) detectors are built upon convolutional neural networks (CNNs), typically categorized into single-stage and two-stage architectures. Single-stage detectors, exemplified by RetinaNet [7], the fully convolutional one-stage object detector (FCOS) [8], and the you only look once (YOLO) family [9–11], are favored due to their high efficiency and real-time capability. Two-stage methods like faster region-based CNN (Faster R-CNN) [12] prioritize localization precision, albeit with a higher computational overhead. Recently, transformer-based approaches, such as the detection transformer (DETR) [13], deformable DETR [14], and real-time detection transformer (RT-DETR) [15], have pushed the boundaries of the field by modeling global dependencies and contextual relationships.

However, most existing detection frameworks are tailored for optical imagery and utilize horizontal bounding boxes (HBB). When directly applied to SAR ship detection, these methods often struggle to accurately capture elongated and oriented vessels, leading to imprecise localization and redundant detections. First, axis-aligned bounding boxes or HBB-based detectors inadequately capture the geometric characteristics of rotated ships, particularly in dense scenarios where target overlapping frequently occurs. Second, small ships with limited pixel coverage are easily confused with background clutter, resulting in missed detections. Third, inherent SAR imaging mechanism often causes dominant scattering centers to shift away from geometric centroids, leading to a significant discrepancy between classification confidence and localization quality. To alleviate these limitations, oriented bounding box (OBB)-based algorithms have received increasing attention in recent years. By introducing additional angular parameters or quadrilateral representations, OBB-based approaches can more precisely localize targets with arbitrary rotations. Nevertheless, many oriented detectors rely on complex architectures or transformer-based designs, which incur heavy computational overhead and limit their suitability for real-time maritime monitoring applications.

Several benchmark datasets have been established to facilitate research on SAR ship detection. Released in 2017, the SAR ship detection dataset (SSDD) [16] serves as one of the earliest public benchmarks in this domain. To overcome resolution constraints, the high-resolution SAR image dataset (HRSID) [17] provides higher spatial resolution and larger image sizes. The high-resolution SAR rotation ship detection dataset (SRSDD-v1.0) [18] extends beyond single class detection by offering multiple ship categories with rotation annotations, enabling more comprehensive evaluation of oriented detection methods.

Motivated by these observations, this work proposes a YOLO-based oriented ship detector (YOSDet) specifically tailored for SAR imagery. While YOLO-based frameworks offer high efficiency, they often require specialized refinements to address unique physical characteristics of SAR. The proposed framework is developed considering several key

challenges in SAR ship detection, including unstable feature representations caused by non-stationary backscattering, inconsistency between localization accuracy and classification confidence for arbitrarily oriented ships, and unreliable confidence estimation under ambiguous scattering conditions. To accommodate these characteristics, YOSDet enhances the backbone with a dynamic aggregation module (DAM) to strengthen feature aggregation under noisy and cluttered SAR imaging conditions. Furthermore, an objective guided detection head (OGDH) is designed to explicitly model the differing requirements of semantic discrimination and geometric regression. In addition, a localization quality estimator (LQE) module based on existing work [19], is incorporated to rectify the discrepancy between classification confidence and prediction precision, ensuring more reliable detection during inference. The core contributions of this research can be outlined as follows.

1. We propose YOSDet, an improved YOLO-based oriented architecture for SAR ship detection that seeks a balance between detection accuracy and inference efficiency.
2. We introduce a dynamic aggregation module (DAM) for robust feature representation and an objective guided detection head (OGDH) with localization quality estimator (LQE) to ensure prediction consistency by mitigating the impact of SAR-specific scattering characteristics.
3. Comprehensive evaluations on SSDD, HRSID, and SRSDD-v1.0 confirm that our YOSDet efficiently and reliably generalizes to both inshore and offshore SAR scenarios.

The subsequent sections of this paper are arranged as follows. Section 2 surveys related work covering generic and oriented object detection, as well as SAR detection methods. The technical details of our proposed method are presented in Section 3. In Section 4, we report the experimental setup and comparative results, followed by a further analysis in Section 5. Finally, Section 6 draws the final conclusions.

2. Related Work

2.1. General Object Detection

The landscape of general object detection has shifted from region-based paradigms toward efficient single-stage frameworks. These two-stage detectors, including faster R-CNN [12] and cascade R-CNN [20], ensure localization precision through region proposal networks (RPN) and iterative refinement. In contrast, single-stage architectures like the RetinaNet [7] and YOLO series [10,11] achieve superior inference speeds by performing classification and regression in an end-to-end process. To enhance feature quality, multi-level feature integration structures such as feature pyramid network (FPN) [21] and bi-directional FPN (BiFPN) [22] have been developed to address scale variations. Furthermore, task-aligned detectors like task-aligned one-stage object detection (TOOD) [23] and dynamic head frameworks [24] have introduced mechanisms to mitigate the spatial semantic misalignment between classification and localization. Recently, transformer-based methods, including DETR [13] and its real-time variant RT-DETR [15], have demonstrated strong global reasoning capabilities through self-attention mechanisms and mixed query selection [25].

2.2. Oriented Bounding Boxes Object Detection

Oriented object detection extends conventional frameworks by predicting rotated bounding boxes to handle targets with arbitrary orientations, which is essential for RS and aerial imagery. In the two-stage paradigm, RoI Transformer [26] and Oriented R-CNN [27] utilize spatial transformations to produce high quality oriented proposals. To achieve orientation alignment, ReDet [28] incorporates rotation-equivariant networks to acquire rotation-invariant features. In the single-stage domain, the refined rotation RetinaNet (R³Det) [29] and single-shot alignment network (S²A-Net) [30] employ feature alignment

modules to harmonize convolutional features with rotated anchors. While these methods achieve impressive accuracy on optical datasets, their performance often fluctuates in scenarios with dense distributions and high-aspect-ratio objects, where the representation of geometric characteristics remains a significant challenge.

2.3. SAR Ship Detection

Detecting vessels in SAR data presents idiosyncratic challenges absent in optical imagery. The presence of speckle noise, sea clutter, and strong reflections from ship structures leads to irregular back-scattering patterns that complicate feature extraction. Researchers have explored various strategies to enhance hierarchical representations and preserve feature information. In [31], a weighted BiFPN is utilized to aggregate multi-scale features, while the YOLO-SRBD [32] introduces shuffle reparameterized blocks combined with dynamic heads to optimize information flow and enhance discriminative capability by reusing redundant features. Additionally, the Quad-FPN [33] designs specialized feature pyramid structures to mitigate scale variations. SMEP-DETR [34] introduces a RT-DETR-based paradigm utilizing parallel dilated convolutions and multi-edge enhancement to aggregate global contextual information for horizontal ship detection without compromising real-time processing speeds.

To handle arbitrary orientations, recent studies have further explored rotation-aware mechanisms and task-specific alignment. The BiFA-YOLO [35] explores multi-level feature aggregation and integrates angle classification to detect objects in various orientations. Further, the rotated balanced feature-aligned network (RBFA-Net) is proposed by [36], introducing a balanced attention FPN (BAFPN) and an anchor-guided feature alignment network (AFAN) with a rotational detection network (RDN) to resolve the misalignment problem between features and OBB anchors. In addition, the WSL-paradigm [37] leverages a weakly supervised strategy to mine pseudo orientations from horizontal annotations, while the multiscale dynamic feature fusion network (MSDFF-Net) [38] employs a dynamic feature fusion block with large-kernel convolutions to balance spatial and channel information in noisy backgrounds. Despite these advancements, achieving a balance between real-time efficiency and localization reliability in complex SAR environments remains an open problem.

Beyond single-frame ship detection, recent research has extended YOLO-based frameworks toward SAR ship tracking. Representative methods such as YOLOShipTracker [39] and two-frame SAR ship tracking framework (TFST) [40] incorporate temporal association mechanisms to enhance cross-frame consistency, while a single-source generalization model for ship detection (SSGNet) [41] focuses on improving environmental adaptability through domain-invariant feature learning. These works primarily address multi-frame consistency and cross-domain generalization and are thus complementary to our focus on reliable single-frame oriented detection under complex SAR scattering conditions.

3. Proposed Method

3.1. Overall Architecture

To harmonize structural complexity with detection accuracy in SAR scenes, this work proposes a YOLO-based end-to-end framework for oriented ship detection, termed YOSDet, as depicted in Figure 1. The YOSDet framework is comprised of three functional networks, including a backbone for hierarchical feature extraction, a neck for feature fusion, and a detection head for prediction.

The data flow originates in the backbone, where an HGStem and a sequence of dynamic aggregation module (DAM) progressively compress spatial resolution while expanding the semantic channel dimension. This process generates three feature levels,

$\{P_3, P_4, P_5\}$, corresponding to downsampling strides of 8, 16, and 32, respectively. To bridge the semantic gap between these scales, the neck network utilizes a path aggregation structure that leverages upsampling and concatenation operations, ensuring that the fused features integrate fine-scale spatial details with high-level information. The final detection is performed by the objective-guided detection head (OGDH) with localization quality estimator (LQE), which decouples classification from regression to resolve specific feature conflicts. This design ensures that the model can independently optimize for semantic discrimination and geometric precision. Meanwhile, the LQE component calibrates the output to ensure high consistency between localization precision and classification confidence, which is particularly critical for the high-aspect-ratio targets typically in SAR imagery.

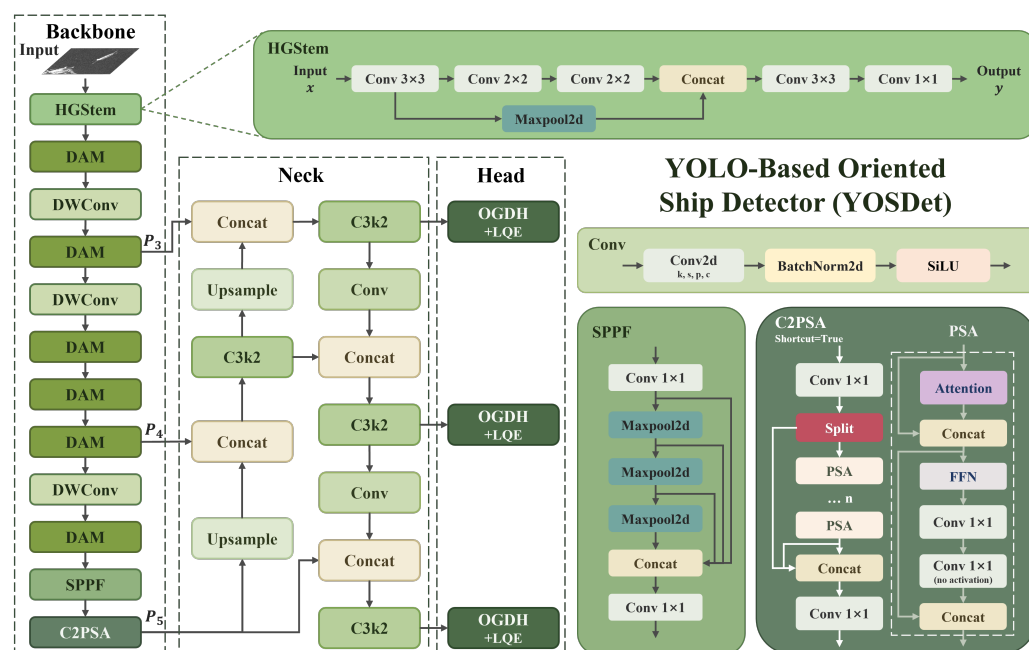


Figure 1. Overall architecture of the proposed framework.

3.2. Feature Extraction Backbone

The backbone of YOSDet is constructed as a hierarchical feature extraction network. It initiates feature extraction via an HGStem module for initial spatial downsampling and low-level feature embedding. The stem output is then processed by a sequence of depthwise convolution (DWConv) layers to progressively extract hierarchical feature representations. A dynamic aggregation module (DAM) is introduced as a backbone aggregation unit to integrate multi-level features. After DAM-based aggregation, high-level features are processed through the spatial pyramid pooling fast (SPPF) block and the convolutional block with parallel spatial attention (C2PSA) as shown in Figure 1. These components jointly form the backbone network, which produces a multi-scale feature hierarchy, $\{P_3, P_4, P_5\}$, to facilitate cross-scale feature integration in the neck.

The DAM is designed as a backbone aggregation unit by leveraging the hierarchical aggregation topology of the HGBlock adopted in RT-DETR [15] and augmenting it with adaptive perception. As depicted in Figure 2, the DAM comprises a cascade of DynamicConv layers, followed by hierarchical feature aggregation with explicit channel reorganization. In DAM, DynamicConv is adopted as the basic feature transformation operator. Unlike conventional convolution with fixed kernel parameters, DynamicConv aggregates multiple parallel convolution kernels dynamically according to the input feature [42]. Considering an input feature map $X \in \mathbb{R}^{C \times H \times W}$ and a set of K parallel convolution kernels $\{W_k\}_{k=1}^K$, the output Y is formulated as

$$Y = \sum_{k=1}^K \pi_k(X) (W_k * X), \quad (1)$$

where $*$ is the convolution operation, and $\pi_k(X)$ represents the aggregation coefficient of the k -th kernel. These coefficients are derived via a lightweight routing function that captures the global context of the input feature.

$$\pi = \text{Softmax}(\text{FC}(\sigma(\text{FC}(\text{GAP}(X))))), \quad (2)$$

where $\text{FC}(\cdot)$ denotes the fully connected layer, $\text{GAP}(\cdot)$ is global average pooling operation, and $\sigma(\cdot)$ denotes the sigmoid linear unit (SiLU) activation operator.

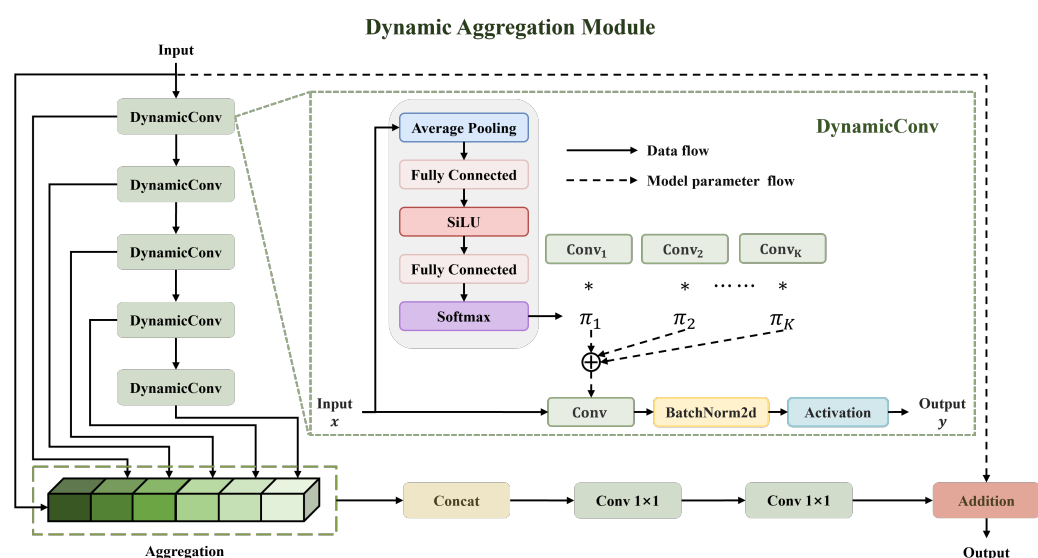


Figure 2. The structure of the dynamic aggregation module, where $*$ denotes the convolution operation.

The integration of DynamicConv within the DAM is motivated by non-stationary backscattering characteristics of SAR imagery. Unlike optical images with stable textures, SAR signals exhibit significant intensity fluctuations depending on incident angles and sea states. This dynamic mechanism allows kernels to adaptively reconfigure their weights based on the specific scattering intensity of the input, which is essential for preserving weak ship signatures while mitigating interference from complex background clutter.

Specifically, within the DAM structure, multiple DynamicConv layers are applied sequentially, where the output from each preceding layer serves as the input for the subsequent one. The original input feature together with all intermediate features are retained and concatenated along the channel dimension, resulting in a progressively aggregated representation that preserves both early structural details and deeper semantic information. Following channel-wise concatenation, the aggregated features are reorganized by two successive 1×1 convolution layers. The primary convolution compresses the concatenated features into a compact intermediate representation to reduce channel redundancy introduced by multi-level aggregation, while the subsequent convolution expands the compressed features to the target output dimension, completing the channel reorganization. When the input and output feature dimensions are identical, a conditional residual connection is applied to further facilitate information reuse and stable optimization. By integrating dynamic convolution with hierarchical feature aggregation and explicit channel reorganization, DAM preserves the structural efficiency of HGBlock-style aggregation while enabling input-adaptive feature extraction under complex sensing conditions.

3.3. Multi-Scale Feature Fusion Neck

To effectively integrate multi-scale representations, YOSDet incorporates a path aggregation network (PAN) to amalgamate hierarchical feature sets $\{P_3, P_4, P_5\}$ extracted from the backbone. Following the neck design of YOLOv11 [10], the proposed framework employs a bidirectional feature fusion strategy consisting of top-down and bottom-up pathways, which helps bridge the semantic gap between different resolution levels. The C3k2 (cross-stage partial with kernel size 2) block serves as a primary component for feature refinement in the neck.

As illustrated in Figure 3, the C3k2 block utilizes a cross-stage partial (CSP) architecture and begins with a 1×1 convolution followed by a split operation that divides the input feature map into two branches. One branch bypasses the transformation layers to preserve original feature information, forming a residual connection, while another branch is fed into n stacked C3k modules. Each C3k module contains two convolutions and a bottleneck structure with 3×3 kernels. These outputs of all branches are subsequently combined through concatenation and a transition convolution. In the top-down pathway, high-level semantic feature maps are upsampled and then concatenated with lower-level backbone features to enhance semantic richness at finer scales. In the bottom-up pathway, the fused feature maps are propagated to deeper levels using 3×3 convolutions, reinforcing fine-grained localization cues. The subsequent detection head benefits from enriched multi-scale representations, as this bidirectional fusion effectively mitigates the interference of complex background clutter and scale variations common in SAR imagery.

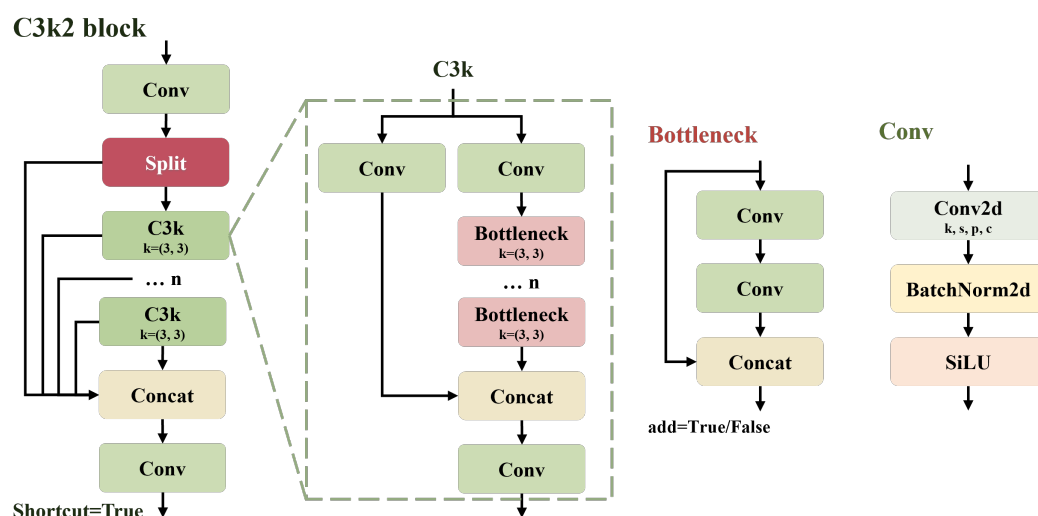


Figure 3. The structure of the C3k2 block with C3k and bottleneck in detail.

3.4. Objective-Guided Detection Head

We introduce an objective-guided detection head (OGDH) to adapt shared features to the individual objectives of classification and localization, as illustrated in Figure 4. Given multi-scale features F_{input} from the neck network, a lightweight shared block composed of two convolutional layers with group normalization (Conv-GN) is first applied. The outputs of these layers are concatenated to form a unified shared representation F_s . Based on F_s , OGDH adopts parallel transformations to derive features specific to classification and regression, denoted as $F_{\text{cls}}^{\text{og}}$ and $F_{\text{reg}}^{\text{og}}$.

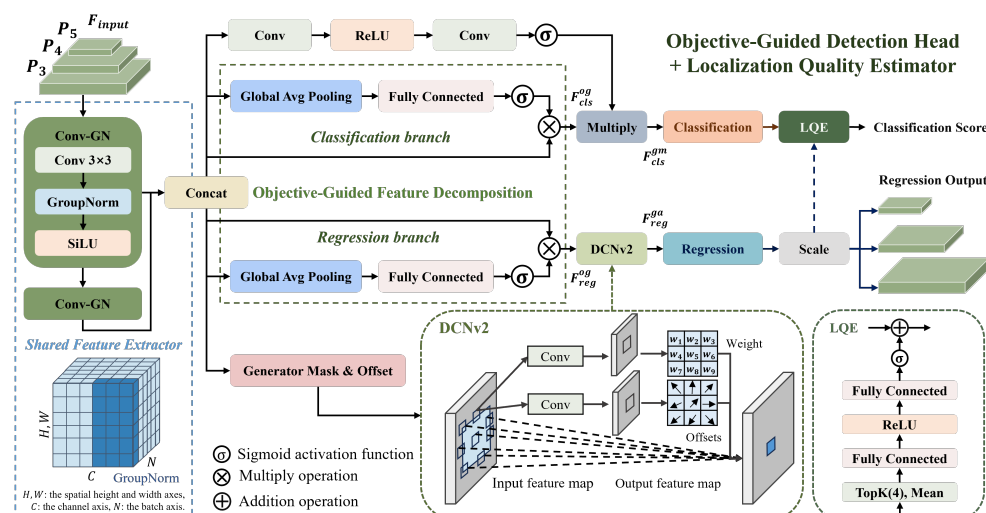


Figure 4. The structure of the objective-guided detection head with localization quality estimator.

Following feature decomposition, objective-guided processing is conducted in the two branches. In the classification branch, a dynamic gating mechanism modulates F_{cls}^{og} via element-wise multiplication, producing a gated feature F_{cls}^{gm} for category prediction. In the regression branch, geometric alignment is achieved using a deformable convolutional network (DCNv2) [43], which enables adaptive spatial sampling to better capture localization information. Specifically, sampling offsets and modulation masks predicted from F_s are applied to F_{reg}^{og} , yielding a geometrically aligned feature F_{reg}^{ga} . The regression branch adopts a distribution-based formulation to predict bounding box parameters (x, y, w, h) . For OBB detection, an additional regression head predicts an angle logit x_θ , which is mapped to the angular range $[-\frac{\pi}{4}, \frac{3\pi}{4}]$ as

$$\theta = (\sigma(x_\theta) - 0.25) \times \pi. \quad (3)$$

The angular regression range $[-\frac{\pi}{4}, \frac{3\pi}{4}]$ spans π radians, which suffices to represent all unique ship orientations owing to the 180° rotational invariance of OBB. Compared with conventional ranges such as $[-\frac{\pi}{2}, \frac{\pi}{2}]$ or $[0, \pi]$, this interval shifts angular discontinuities away from the horizontal and vertical axes to diagonal directions. This design choice is intended to alleviate the boundary sensitivity for ship orientations near the principal axes during angle regression, facilitating stable optimization.

Localization Quality Estimator

The adoption of LQE is necessitated by spatial misalignment between dominant scattering centers and the target's geometric centroid in SAR imagery. Strong multi-path reflections often concentrate on local metallic structures, causing a high classification confidence to be assigned to proposals with poor geometric alignment. LQE addresses this by extracting statistical uncertainty from the regression distribution to calibrate classification scores, ensuring consistency between detection confidence and actual localization precision.

Under this framework, the regression branch predicts discrete probability distributions for the four bounding box boundaries, defined as $\mathcal{W} = \{l, r, t, b\}$. The corresponding discrete probability distribution for each boundary is denoted as

$$P^w = [P^w(y_0), P^w(y_1), \dots, P^w(y_N)], \quad w \in \mathcal{W}, \quad (4)$$

where y_i represents discretized offset bins.

Based on the observation that the sharpness of a distribution reflects localization certainty, LQE extracts statistical features from each boundary distribution by selecting

the top- k highest probabilities together with their mean value. The resulting localization descriptor is formulated as

$$\mathbf{F} = \text{Concat}(\{\text{Topkm}(P^w) \mid w \in \mathcal{W}\}), \quad (5)$$

where $\mathbf{F} \in \mathbb{R}^{4(k+1)}$, and $\text{Topkm}(\cdot)$ represents the integrated procedure of top- k selection and mean averaging.

The extracted feature $\mathbf{F}$ is subsequently fed into a lightweight multilayer perceptron to predict a localization quality correction term,

$$\Delta I = \mathcal{F}(\mathbf{F}) = \mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{F}), \quad (6)$$

where $\delta(\cdot)$ denotes the ReLU activation function. To ensure stable optimization, the output layer is initialized to zero, such that the LQE initially performs an identity mapping and gradually learns effective confidence calibration during training. Instead of directly rescaling classification scores, LQE performs residual calibration on the classification logits,

$$\tilde{C} = C + \Delta I, \quad (7)$$

where C denotes the original classification logits, and $\tilde{C}$ represents the LQE-calibrated logits.

The LQE branch is trained in an implicit manner without introducing explicit supervision on localization quality. Since the distribution sharpness already encapsulates localization certainty, the correction term can be optimized indirectly through the classification loss. This enables the model to autonomously learn confidence calibration consistent with localization reliability, rendering additional quality labels redundant.

3.5. Loss Function

The overall training objective is formulated as a weighted sum of three components, including a classification loss (L_{cls}), an OBB regression loss (L_{obb}), and a distribution focal loss (L_{dfl}). The total loss is defined as

$$L = \lambda_{cls} L_{cls} + \lambda_{obb} L_{obb} + \lambda_{dfl} L_{dfl}, \quad (8)$$

where λ_{cls} , λ_{obb} , and λ_{dfl} are trade-off coefficients used to regulate the relative importance of each loss component.

The classification loss is computed using binary cross-entropy with logits, where the LQE-calibrated logits $\tilde{C}$ are supervised by ground truth (GT) labels,

$$L_{cls} = \frac{1}{N} \sum_i \left(-y_i \log(\sigma(\tilde{C}_i)) - (1 - y_i) \log(1 - \sigma(\tilde{C}_i)) \right), \quad (9)$$

where y_i represents the GT label for the i -th positive instance, and $\sigma(\cdot)$ and N signify the sigmoid operator and the total count of positive samples, respectively.

For OBB regression, we adopt the probabilistic intersection over union (ProbIoU) loss, which models each bounding box as a 2D Gaussian distribution and measures their similarity in a probabilistic principle [44]. Given the predicted and GT distributions $p(x)$ and $q(x)$, the Bhattacharyya coefficient and distance are formulated as follows.

$$B_C(p, q) = \int_{\mathbb{R}^2} \sqrt{p(x)q(x)} dx, \quad (10)$$

$$H_D(p, q) = \sqrt{1 - B_C(p, q)}. \quad (11)$$

The regression loss is formulated as

$$L_{obb} = H_D(p, q) = 1 - \text{ProbIoU}(p, q) \in [0, 1]. \quad (12)$$

The distribution focal loss (DFL) is further employed to refine the regression of bounding box parameters (x, y, w, h) by supervising predicted discrete distributions. Following the formulation in generalized focal loss, DFL linearly interpolates two adjacent bins surrounding the continuous target value,

$$L_{dfl} = -((y_{i+1} - y) \log(S_i) + (y - y_i) \log(S_{i+1})), \quad (13)$$

where y_i and y_{i+1} are the consecutive integer labels satisfying $y_i < y < y_{i+1}$, and S_i, S_{i+1} are the predicted probabilities from a Softmax layer, $S_i = \frac{y_{i+1}-y}{y_{i+1}-y_i}$ and $S_{i+1} = \frac{y-y_i}{y_{i+1}-y_i}$.

4. Experiments and Results

4.1. Dataset

We evaluate the proposed YOSDet on three widely recognized benchmarks for oriented SAR ship detection: SSDD+ [16], HRSID [17], and SRSDD-v1.0 [18]. Table 1 provides a comprehensive overview of the statistical characteristics for these datasets.

Table 1. Statistics of SSDD+, HRSID, and SRSDD-v1.0.

Details	SSDD (SSDD+)	HRSID	SRSDD-v1.0
Sources	RadarSat-2, TerraSAR-X, Sentinel-1	Sentinel-1, TerraSAR-X, TanDEM-X	Gaofen-3
Polarization	HH, HV, VV, VH	HH, HV, VV	HH, VV
Resolution (m)	1~15	0.5, 1, 3	1
Dimensions (pixel)	190~668	800 × 800	1024 × 1024
Images/Instances	1160/2456	5604/16,951	666/2884
Annotations	HBB, OBB, Polygon	Polygon	OBB
Categories	1	1	6

The SSDD+ [16], which inaugurated public SAR ship detection research, comprises 1160 images and 2456 annotated ships. The data are sourced from multiple SAR sensors, such as RadarSat-2 and Sentinel-1, covering diverse sea states and imaging conditions. Following common practice, the dataset is split into training and test subsets using an 8:2 ratio, assigning 928 training images and 232 test images. The HRSID [17] is a high-resolution benchmark comprising 5604 SAR images and 16,951 ship targets, specifically designated for detection and segmentation tasks. In this work, OBB labels are derived from the native instance segmentation masks by fitting the minimum area rotated bounding box. The dataset is divided into training and testing sets using a 65%:35% split, yielding 3642 training images and 1962 test images. The SRSDD-v1.0 [18] integrates 666 Gaofen-3 SAR images, featuring 2884 ship targets categorized into six types: ore-oil, bulk-cargo, fishing, law enforcement, dredger, and container ships. Approximately 63.1% of the images cover complex inshore scenes with intense background clutter and scattering interference. Following the official data protocol, the imagery is partitioned into a training set of 532 images and a test set of 134 images.

4.2. Implementation Details and Evaluation Metrics

Our experimental framework is built upon PyTorch 2.1.0 and conducted under identical hardware conditions. We standardize the input resolutions for SSDD+ and HRSID to 512×512 and 800×800 pixels, respectively, while the SRSDD-v1.0 images are scaled to 640×640 . Training is conducted over 300 epochs, utilizing a batch size of 4 for the

former two datasets and 2 for the latter. The optimization process is driven by a stochastic gradient descent (SGD) algorithm, configured with a 0.01 learning rate, 0.937 momentum, and 5×10^{-4} weight decay. For comparative analysis, various SOTA oriented detectors are rigorously assessed against the same configuration settings via MMRotate platform [45].

To evaluate the oriented object detection performance, several standard metrics are employed, including precision (P), recall (R), $F1$ score, and mean average precision (mAP). The $F1$ -score is computed as the harmonic mean of precision and recall. The inference efficiency is measured by frames per second (FPS). These core evaluation criteria are formulated as follows.

$$\text{Precision } (P) = \frac{TP}{TP + FP}, \quad (14)$$

$$\text{Recall } (R) = \frac{TP}{TP + FN}, \quad (15)$$

$$\text{F1-score} = \frac{2 \cdot P \cdot R}{P + R}, \quad (16)$$

where TP represents correctly detected instances (true positives), FP denotes incorrect detections (false positives), and FN indicates missed objects (false negatives).

The AP is determined by integrating over the precision–recall curve, and the mAP represents the mean AP across all N object categories:

$$AP = \int_0^1 P(R) dR, \quad mAP = \frac{1}{N} \sum_{n=1}^N AP_n. \quad (17)$$

The FPS metric quantifies the model's processing speed and is calculated by

$$FPS = \frac{s}{T}, \quad (18)$$

where s and T represent the number of samples and the corresponding processing time, respectively.

4.3. Comparisons of Performance

To evaluate the performance of YOSDet, we compare it against several SOTA oriented object detectors, focusing on methods designed for OBB tasks to ensure a consistent evaluation protocol under oriented representation and metrics. The comparison includes classical two-stage methods like RoI Transformer [26], Oriented R-CNN [27], and ReDet [28] and representative single-stage detectors including R³Det [29], S²A-Net [30], and Rotated FCOS [8], as well as recent lightweight YOLO-based models, YOLOv11-OBB [10] and YOLOv13-OBB [11]. Our evaluation prioritizes detection accuracy, robustness across different maritime scenes, and computational efficiency. As summarized in Tables 2 and 3, YOSDet achieves a superior trade-off between these performance metrics and model complexity.

On the SSDD+ dataset, YOSDet achieves a peak mAP of 96.8%, which is a 1.9%p improvement over the second-best method, YOLOv13-OBB. In challenging inshore scenes, where heavy land clutter and complex electromagnetic scattering typically degrade performance, YOSDet maintains a robust mAP of 92.1% and an $F1$ score of 90.9%, surpassing YOLOv13-OBB by 8.3%p and 10.3%p, respectively. This performance indicates that the proposed method effectively suppresses background interference and maintains high localization accuracy for targets in dense arrangements.

Table 2. Quantitative assessment of detection efficacy on SSDD+ across diverse scenarios (%).

Method	Param (M)	FLOPs (G)	Entire Scene				Inshore Scene				Offshore Scene				FPS
			P	R	mAP	F1	P	R	mAP	F1	P	R	mAP	F1	
R-Faster-RCNN [12]	41.12	63.25	87.8	78.9	78.0	83.1	66.2	60.5	53.7	63.2	91.4	93.3	89.2	92.3	88.1
Gliding Vertex [46]	41.13	63.25	89.6	80.8	80.2	85.0	78.4	57.0	57.9	66.0	92.1	93.6	90.0	92.8	86.1
RoI Transformer [26]	55.03	77.15	94.3	84.4	88.0	89.1	86.4	66.3	68.8	75.0	96.4	94.1	90.6	95.3	61.6
Oriented R-CNN [27]	41.35	63.28	92.3	83.5	87.2	87.7	76.9	65.7	66.9	70.8	97.2	92.8	90.4	94.9	52.8
R-RetinaNet [7]	36.13	52.39	84.0	79.9	78.4	81.9	55.7	51.2	45.6	53.3	95.1	93.9	90.0	94.5	132.7
ReDet [28]	31.54	40.88	96.2	89.2	90.3	92.6	88.4	70.9	71.3	78.7	97.9	98.9	90.9	98.4	34.5
R ³ Det [29]	41.58	82.17	97.0	84.2	87.5	90.2	90.6	61.6	67.5	73.4	98.4	96.3	90.8	97.3	45.6
S ² A-Net [30]	38.54	49.05	92.0	90.7	89.9	91.3	82.1	77.3	76.3	79.6	96.3	96.8	90.7	96.5	90.6
R-ATSS [47]	36.01	51.79	93.9	87.0	88.5	90.3	82.6	66.3	66.4	73.5	97.6	97.9	90.7	97.7	141.9
Rotated FCOS [8]	31.89	51.55	87.8	80.0	79.4	83.7	63.4	56.4	53.3	59.7	93.6	93.6	89.6	93.6	144.9
Rotated RepPoints [48]	36.60	48.56	85.6	77.5	77.5	81.3	62.9	48.3	48.7	54.6	89.4	94.7	88.5	91.9	112.0
Oriented RepPoints [49]	36.60	48.56	95.1	85.0	89.0	89.7	84.9	62.2	70.6	71.8	97.3	96.8	90.8	97.1	76.5
SASM RepPoints [50]	36.60	48.56	90.9	85.5	87.8	88.1	86.2	61.6	69.6	71.9	97.2	94.1	90.7	95.7	111.6
YOLOv11-OB [10]	2.65	4.20	97.2	90.3	94.0	93.6	93.4	74.4	82.3	82.9	98.1	98.1	98.6	98.1	120.5
YOLOv13-OB [11]	2.52	4.10	92.7	92.3	94.9	92.5	81.6	79.6	83.8	80.6	98.1	98.1	98.6	98.1	117.6
WSL paradigm [37]	55.84	-	-	89.7	87.3	-	-	73.3	67.5	-	-	95.0	93.9	-	19.7
YOSDet	2.15	5.00	97.4	94.7	96.8	96.0	94.9	87.2	92.1	90.9	98.1	98.7	98.6	98.4	70.4

Table 3. Quantitative assessment of detection efficacy on HRSID across diverse scenarios (%).

Method	Param (M)	FLOPs (G)	Entire Scene				Inshore Scene				Offshore Scene				FPS
			P	R	mAP	F1	P	R	mAP	F1	P	R	mAP	F1	
R-Faster-RCNN [12]	41.12	134.38	85.9	64.9	67.8	74.0	57.1	45.5	42.8	50.6	94.1	90.9	88.2	92.5	51.9
Gliding Vertex [46]	41.13	134.39	85.2	66.8	69.5	74.9	61.3	44.3	43.9	51.5	95.6	90.6	90.3	93.1	55.3
RoI Transformer [26]	55.03	148.28	86.7	72.7	76.6	79.1	69.3	52.0	53.0	59.4	97.2	93.4	90.7	95.2	42.3
Oriented R-CNN [27]	41.35	134.46	89.2	73.2	78.1	80.4	69.7	55.4	55.5	61.7	97.0	94.5	90.7	95.7	65.7
R-RetinaNet [7]	36.13	128.09	79.2	65.3	66.7	71.6	53.5	40.7	37.7	46.3	92.1	92.0	88.5	92.1	61.3
ReDet [28]	31.54	59.74	88.3	77.5	79.6	82.6	75.7	58.4	62.1	65.9	97.4	95.5	90.7	96.5	31.6
R ³ Det [29]	41.58	200.92	90.3	71.5	76.8	79.8	70.3	53.7	54.7	60.9	96.2	93.8	90.7	94.9	42.4
S ² A-Net [30]	38.54	119.92	91.0	76.0	79.6	82.8	75.5	58.8	61.7	66.1	98.0	95.0	90.8	96.5	48.9
R-ATSS [47]	36.01	126.62	86.1	70.8	74.5	77.7	68.1	49.9	51.0	57.6	95.7	92.8	89.9	94.3	63.7
Rotated FCOS [8]	31.89	125.98	82.0	69.3	74.1	75.1	60.4	51.1	49.1	55.3	92.7	91.8	89.8	92.3	72.3
Rotated RepPoints [48]	36.60	118.72	76.7	70.8	71.7	73.6	59.5	51.2	49.5	55.1	90.7	92.7	88.6	91.7	59.6
Oriented RepPoints [49]	36.60	118.72	85.4	79.7	79.4	82.5	71.9	63.1	63.8	67.2	97.2	94.6	90.6	95.9	72.1
SASM RepPoints [50]	36.60	118.72	88.1	71.5	77.4	78.9	69.6	54.5	57.6	61.1	95.1	91.9	90.4	93.5	55.4
YOLOv11-OB [10]	2.65	10.20	89.9	79.1	86.3	84.2	77.4	62.7	69.4	69.3	97.4	96.2	97.5	96.8	108.7
YOLOv13-OB [11]	2.52	10.00	89.3	80.1	86.8	84.4	76.8	65.8	71.9	70.9	97.6	94.7	96.7	96.1	109.9
WSL paradigm [37]	55.84	-	-	85.0	81.5	-	-	78.9	71.6	-	-	96.2	94.8	-	16.9
MSDF-Net [38]	8.94	-	83.6	88.1	83.1	85.8	69.7	75.5	70.1	72.5	98.4	98.0	91.9	98.2	-
YOSDet	2.15	12.30	90.2	82.2	88.5	86.0	80.1	68.5	75.4	73.9	96.9	96.0	97.1	96.5	108.7

On the HRSID dataset, YOSDet obtains an *mAP* of 88.5% with a parameter count of only 2.15 M. Compared with YOLOv13-OB, YOSDet yields a 1.7% improvement in *mAP* while operating at a comparable inference speed. The notable enhancement in recall, which reaches 82.2% for the entire scene compared to 80.1% for YOLOv13-OB, indicates that YOSDet effectively mitigates missed detections for small and densely distributed ships, which remain a major challenge in high-resolution SAR imagery.

The performance comparison across various ship classes on SRSD-v1.0 is detailed in Table 4. YOSDet outperforms all compared methods with the highest *mAP* of 67.3%, exceeding the second-best TIAR-SAR by 3.4%. The results highlight that YOSDet excels in identifying categories with high geometric complexity and irregular scattering patterns, particularly fishing boats and container ships. In the fishing boat category, YOSDet achieves 46.2% *AP*, which is 4.7% higher than the second-best RBFA-Net. Similarly, for container ships, YOSDet reaches 85.9% *AP*, leading the second-best YOLOv11-OB by 7.0%. The consistent performance across all categories demonstrates the generalizability of the model in practical diverse ship categories and complex maritime environments.

In addition to the detection accuracy, computational efficiency is paramount for edge deployment. While YOSDet increases the FLOPs from 4.20 G to 5.00 G on SSDD+ and from 10.20 G to 12.30 G on HRSID, this marginal overhead is a deliberate strategic exchange that yields a significant 18.9% reduction in total parameters. The proposed modules are designed as structural replacements for the original backbone blocks and detection head

rather than additive plug-in modules, allowing most computations to reuse existing feature resolutions and operations. With only 2.15 M parameters, representing approximately 5.2% of those used by Oriented R-CNN, and an inference speed of up to 108.7 FPS on the HRSID, YOSDet attains an optimal equilibrium between accuracy and efficiency. This high inference efficiency, combined with its competitive precision, confirms that YOSDet is well-suited for deployment in resource-constrained SAR ship detection applications.

Table 4. Quantitative assessment of detection efficacy on SRSDD-v1.0 across diverse categories (%).

Method	Ore–Oil	Bulk-Cargo	Fishing	Law Enf.	Dredger	Container	mAP
R-Faster-RCNN [12]	54.6	45.9	21.6	9.1	78.2	72.2	46.9
Gliding Vertex [46]	43.7	44.8	25.8	3.9	74.5	76.6	44.9
RoI Transformer [26]	64.8	49.4	24.3	13.6	70.7	71.0	49.0
Oriented R-CNN [27]	61.8	57.6	33.4	27.3	78.7	78.1	56.1
R-RetinaNet [7]	39.1	30.0	21.5	0.6	56.5	51.8	33.2
ReDet [28]	59.9	46.6	25.5	27.3	71.9	77.2	51.4
R ³ Det [29]	54.5	51.5	25.4	28.2	77.9	82.9	53.4
S ² A-Net [30]	63.4	45.0	30.5	20.9	74.4	77.3	51.9
R-ATSS [47]	52.6	44.1	22.3	54.5	76.0	81.6	55.2
Rotated FCOS [8]	57.5	42.8	25.4	31.3	80.8	71.7	51.6
Rotated RepPoints [48]	40.5	36.6	21.9	0.1	78.0	63.2	40.1
Oriented RepPoints [49]	61.1	49.4	48.8	17.7	80.3	81.3	56.4
SASM RepPoints [50]	59.4	43.5	29.8	1.0	75.4	74.4	47.3
YOLOv11-OB [10]	48.4	49.7	30.0	44.9	70.5	78.9	53.7
YOLOv13-OB [11]	51.3	51.1	14.6	62.1	84.2	76.4	56.6
RBFA-Net [36]	59.4	57.4	41.5	73.5	77.2	71.6	63.4
TIAR-SAR [51]	55.7	69.3	33.1	100.0	70.8	54.5	63.9
YOSDet	49.4	60.2	46.2	78.9	83.4	85.9	67.3

To further evaluate the detection performance under challenging SAR conditions, Figures 5 and 6 present qualitative visual results on SSDD+ and HRSID. For SSDD+ (Figure 5), YOSDet successfully distinguishes ships from complex land clutter (a) and resolves densely clustered targets (b) using orientation-aware bounding boxes, effectively suppressing clutter-induced false alarms. In inshore scenes with small ships and coastal structures (c) and offshore low-SNR environments with multiple small targets (d), the model demonstrates robust detection, capturing weak or closely spaced ships that baseline detectors often miss. On HRSID (Figure 6), YOSDet maintains precise localization across varied inshore scenarios (a–d) and shows strong scale adaptability for both extremely small and giant vessels in offshore scenes (e,f). Overall, these visualizations confirm YOSDet’s ability to mitigate both missed detections and false alarms across diverse maritime environments.

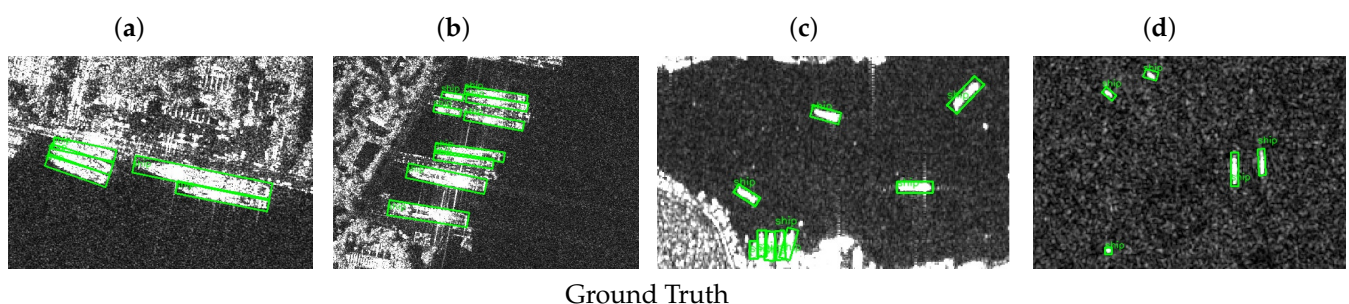
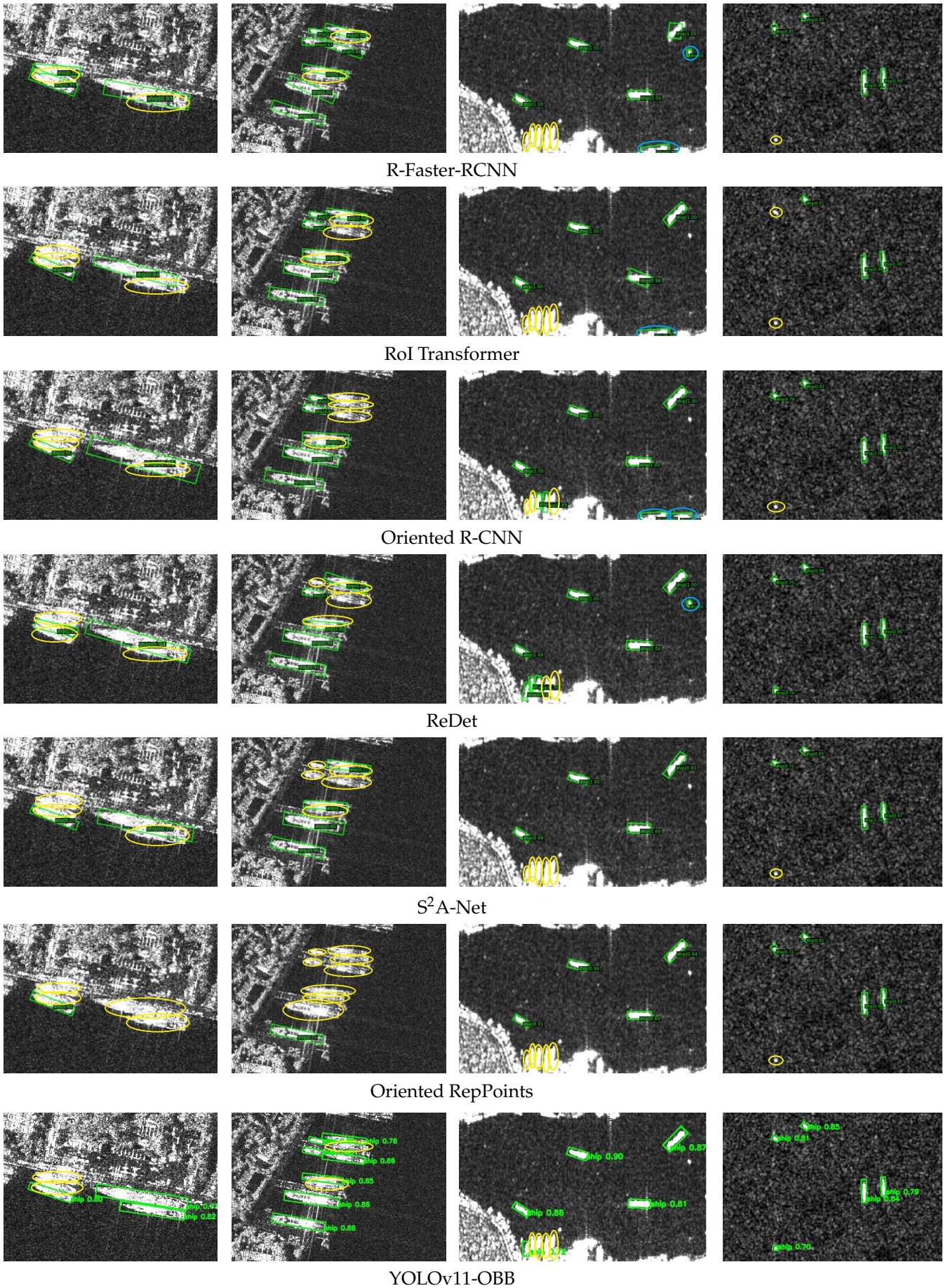


Figure 5. Cont.

Figure 5. *Cont.*

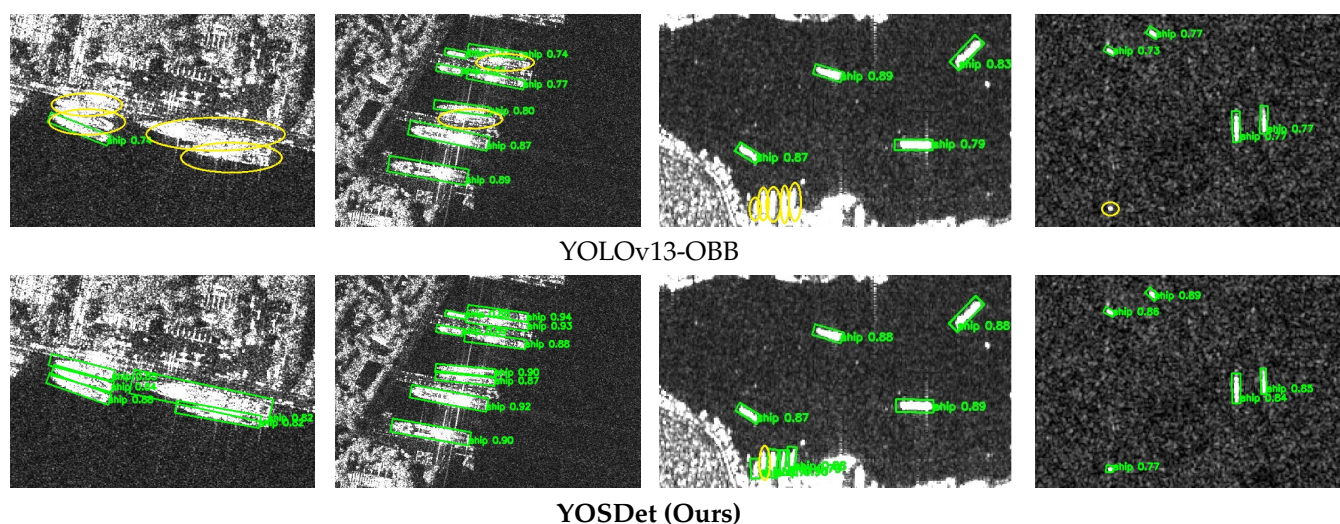


Figure 5. Qualitative detection results on SSDD+. (a) Inshore scene with complex land clutter, (b) inshore scene with densely clustered ships, (c) inshore scene with small ships and coastal structures, and (d) offshore scene with multiple small targets in a low signal-to-noise ratio (SNR) scenario. Green rectangles represent predicted OBb, and yellow and blue circles indicate missed detections and false alarms, respectively.

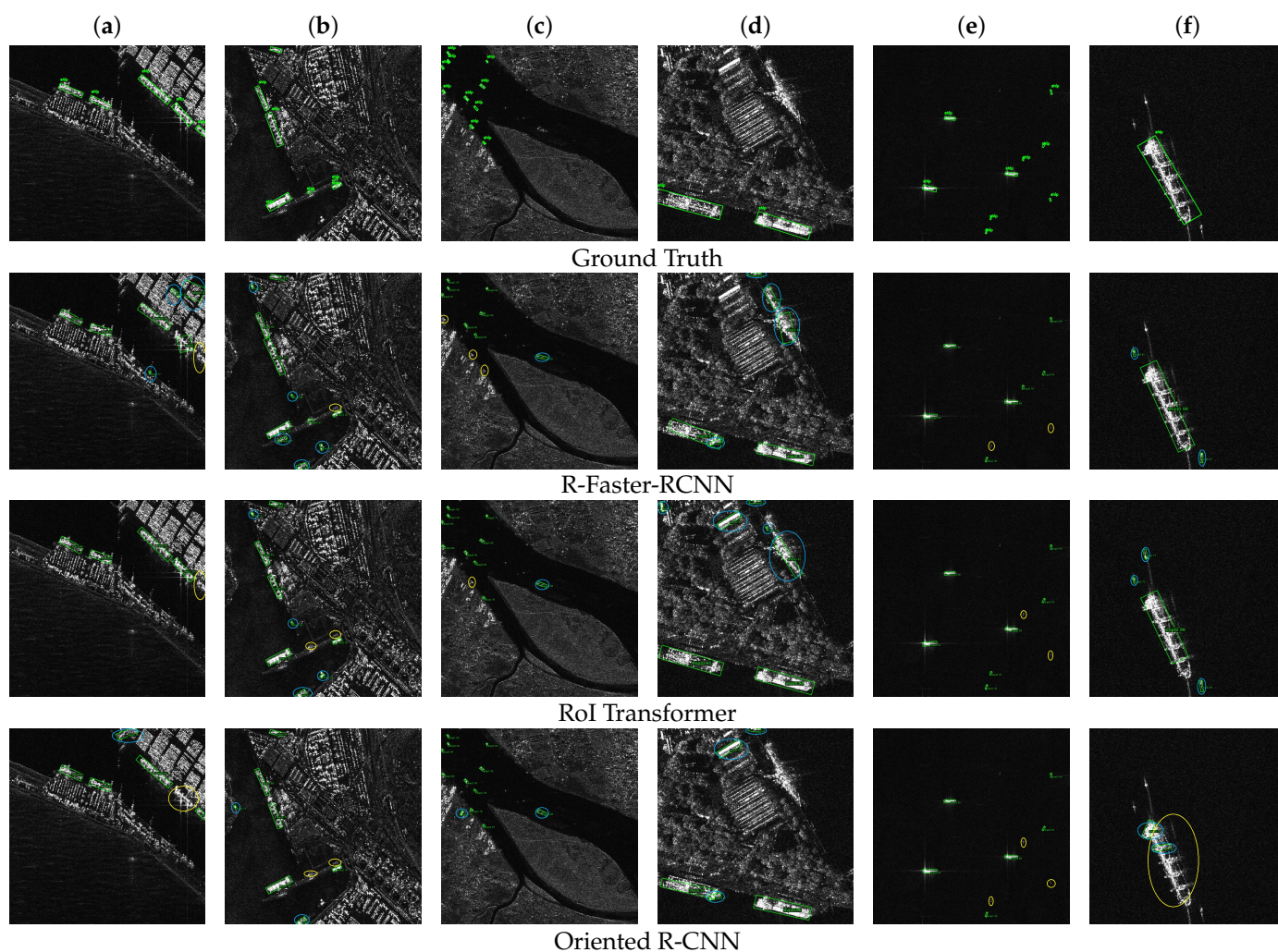


Figure 6. Cont.

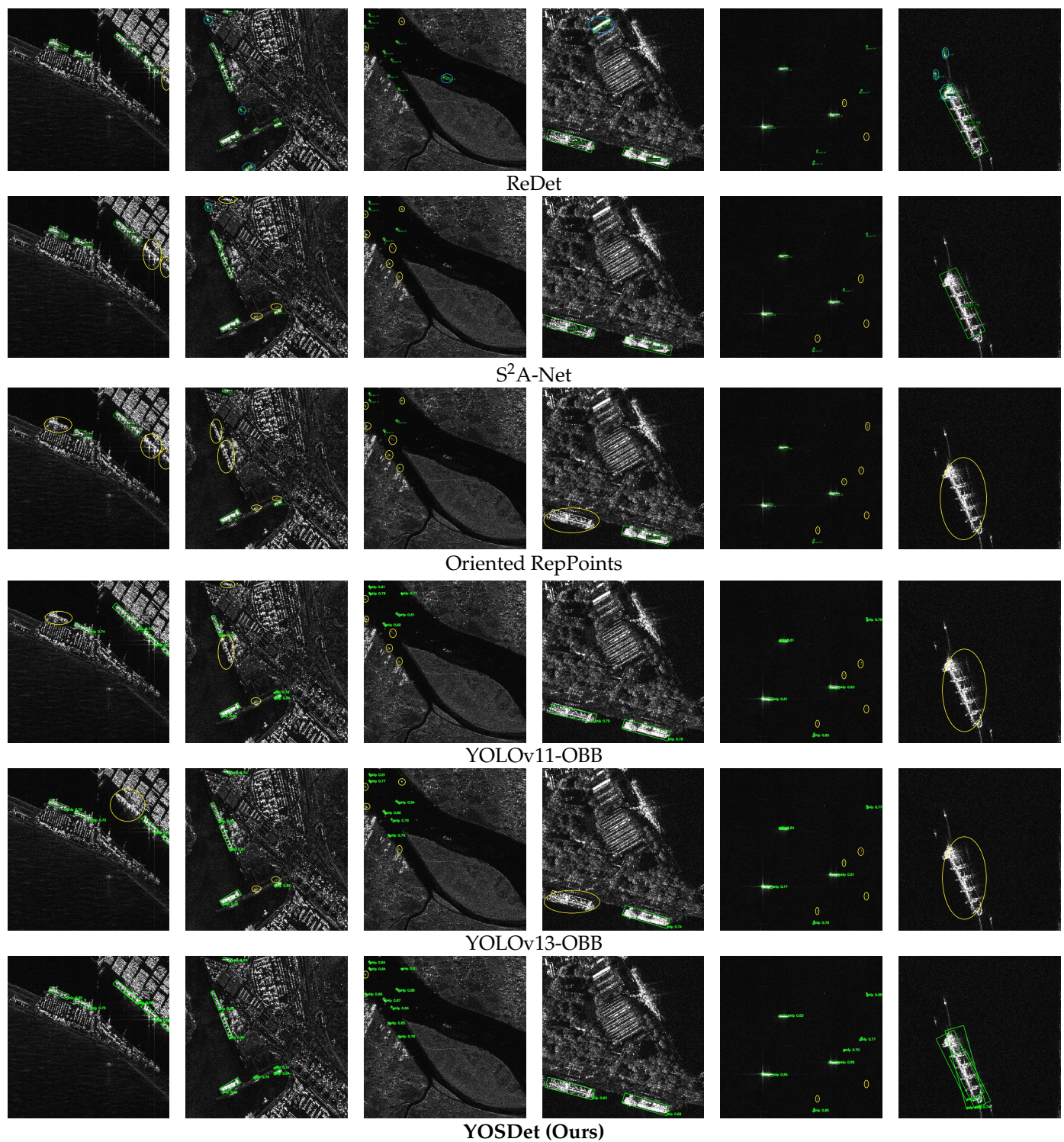


Figure 6. Qualitative detection results on HRSID. (a–d) Inshore scenes and (e,f) offshore scenes. (a) Moored ships along a wharf, (b) ships of various sizes within complex backgrounds, (c) multiple small vessels at an estuary, (d) large ships near the shoreline, (e) extremely small ships in the open sea, and (f) a giant vessel. Green rectangles represent the predicted OBBs, and yellow and blue circles indicate missed detections and false alarms, respectively.

4.4. Ablation Study

Ablation experiments are performed on SSDD+, HRSID, and SRSD-1.0 to evaluate the contribution of each proposed component. The detailed analysis is summarized in Table 5. The OGDH facilitates task coordination and structural optimization. By replacing

the standard OBB detection head with our proposed OGDH as the Row (3), the model parameters significantly decrease from 2.65 M to 2.24 M. This reduction indicates that our objective-guided design is more parameter efficient than the baseline head. Despite the leaner architecture, the *mAP* on the complex SRSDD-v1.0 dataset improves from 53.7% to 59.4%, demonstrating that OGDH effectively resolves the inherent conflicts between classification and localization through superior task decomposition.

Similarly, the DAM substitutes original feature extraction units to strengthen the backbone's representational ability, particularly in suppressing heavy background clutter in SAR images. This substitution is evidenced by the consistent boost in recall (*R*), such as the improvement from 90.3% to 92.1% on SSDD+. By prioritizing salient ship related regions rather than redundant global information, DAM ensures that the model captures potential targets more effectively during the feature encoding stage.

The integration of the LQE module provides a critical calibration for OBB predictions. While LQE addition on the baseline model brings negligible parameters, the version adapted for the OGDH framework introduces a slight increment (+0.1M) to match the higher feature dimensionality maintained within the OGDH structure. This refined configuration yields substantial returns in localization precision; the full proposed framework (Row 6) achieves a peak *mAP* of 67.3% and an *mAP*_{50:95} of 34.4% on SRSDD-v1.0. Consequently, integrating these modules yields an improvement of 13.6% *mAP* over the baseline, while requiring only 2.15 M parameters.

Table 5. Ablation study of different module combinations on SSDD+, HRSID, and SRSDD-v1.0.

	Component			Params (M)	SSDD+				HRSID				SRSDD-v1.0			
	DAM	OGDH	LQE		P	R	mAP	mAP _{50:95}	P	R	mAP	mAP _{50:95}	P	R	mAP	mAP _{50:95}
(1)	—	—	—	2.65	97.2	90.3	94.0	50.9	89.8	79.1	86.3	46.9	54.7	48.2	53.7	27.1
(2)	+	—	—	2.47	95.3	92.1	94.3	51.8	89.0	80.9	85.8	47.2	63.2	53.9	56.3	31.2
(3)	—	+	—	2.24	96.0	92.7	95.5	48.8	91.7	79.2	86.9	47.2	54.5	64.8	59.4	32.7
(4)	—	—	+	2.66	97.3	91.6	95.1	52.1	90.4	80.7	87.9	47.2	59.6	60.4	58.3	30.9
(5)	+	+	—	2.05	96.7	91.6	94.8	52.2	89.2	80.9	87.1	46.5	65.2	61.1	64.8	33.2
(6)	+	+	+	2.15	97.4	94.7	96.8	54.4	90.2	82.2	88.5	49.1	64.0	66.8	67.3	34.4

To analyze the sensitivity of key hyperparameters, we conduct a detailed ablation study on the DAM and OGDH modules as summarized in Table 6. For the DAM module, different numbers of dynamic convolution kernels ($k \in \{2, 4, 8\}$) are evaluated. Increasing k from 2 to 4 leads to a substantial 1.7% improvement in *mAP*_{50:95} and improves performance in inshore scenarios, indicating enhanced localization robustness under complex clutter conditions. Although $k = 2$ exhibits a marginal advantage in simple offshore scenes, it lacks competitiveness in complex areas. Further increasing k to 8 does not bring additional performance gains while introducing extra parameters. For the OGDH module, we investigate the impact of the group normalization parameter ($g \in \{8, 16, 32\}$). While $g = 8$ shows slight gains in offshore scenarios, $g = 16$ achieves a superior *mAP*_{in} of 92.1%, outperforming the $g = 8$ configuration by 2.6%. In contrast, setting $g = 32$ degrades both *P* and *R*, suggesting that overly fine-grained normalization may disrupt stable feature statistics. These observations suggest that the combination of $k = 4$ and $g = 16$ provides a more robust and stable configuration across diverse SAR maritime scenarios. Notably, the FLOPs remain constant at 5.00 G across all configurations of k and g , indicating that hyperparameter tuning does not introduce additional computational overhead or latency.

As illustrated in Figure 7, the proposed LQE improves the consistency between classification confidence and localization quality. Compared with the baseline model, the confidence scores calibrated by LQE exhibit a clearer positive correlation with IoU and are closer to the ideal alignment line. This improvement is further reflected by a reduction in

the mean absolute error (MAE) from 0.138 to 0.130, indicating that the calibrated confidence better reflects localization reliability.

Table 6. Sensitivity analysis of hyperparameters in DAM and OGDH modules on SSDD+ across diverse scenarios.

(a) Impact of dynamic convolution kernel number k .										
	k	P	R	mAP	mAP _{50:95}	mAP _{in}	mAP _{off}	Params	GFLOPs	FPS
DAM	2	96.2	92.5	96.7	52.7	91.4	98.7	1.98	5.00	69.9
	4	97.4	94.7	96.8	54.4	92.1	98.6	2.15	5.00	70.4
	8	96.6	93.0	96.6	54.2	91.3	98.6	2.51	5.00	70.4
(b) Impact of group normalization number g .										
	g	P	R	mAP	mAP _{50:95}	mAP _{in}	mAP _{off}	Params	GFLOPs	FPS
OGDH	8	94.1	93.8	96.4	53.3	89.5	98.9	2.15	5.00	69.9
	16	97.4	94.7	96.8	54.4	92.1	98.6	2.15	5.00	70.4
	32	96.5	91.6	95.9	53.1	89.3	98.3	2.15	5.00	69.4

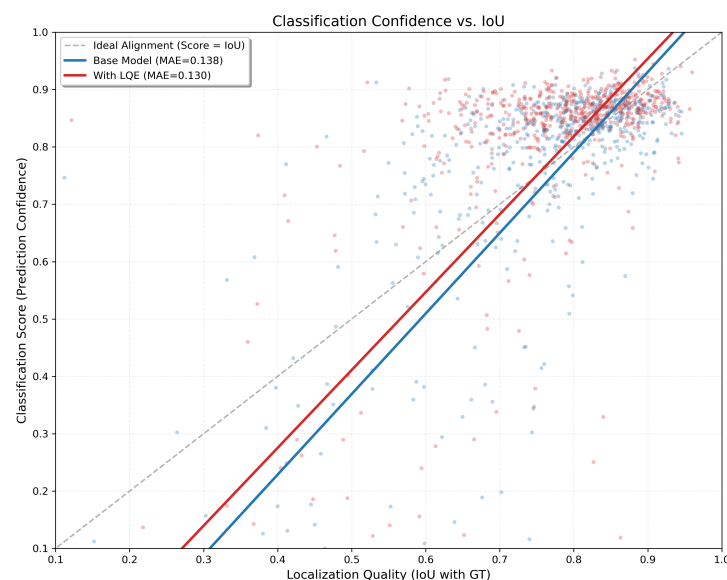


Figure 7. Comparison of confidence–IoU correlation with and without LQE. The dashed line denotes the ideal $Score = IoU$ alignment.

5. Discussion

These experimental results demonstrate that, for SAR ship detection, optimizing the feature aggregation and task alignment is more important than simply increasing the network depth. YOSDet maintains high precision in nearshore scenarios (92.1% mAP on SSDD+ and 75.4% mAP on HRSID), where baseline models often struggle with land clutter. This robustness indicates that adaptive feature aggregation plays an important role in mitigating the influence of complex coastal scattering, rather than relying solely on deeper network structures. Unlike standard convolutions with fixed weights, the DAM adjusts its kernels dynamically based on input backscattering patterns, which provides a flexible mechanism to adapt feature responses under heterogeneous SAR imaging conditions, enabling the model to emphasize structured hull scattering while suppressing random reflections from surrounding infrastructure.

On the SRSD+ dataset, container ship detection achieves 85.9% AP , further demonstrating the synergy between OGDH and LQE. In SAR imagery, dominant scattering peaks often originate from localized structures such as the bridge, which may not align

with the geometric center of the target. Standard coupled heads often produce skewed bounding boxes in these cases, whereas OGDH resolves this by decoupling boundary regression from classification features. LQE complements this approach by calibrating detection scores based on regression sharpness, ensuring that high-aspect-ratio vessels remain tightly bounded even under irregular scattering or partial occlusion.

Despite the overall performance gains, the detection accuracy for fishing boats on SRSDD-v1.0 remains limited (46.2% AP). This is mainly due to the extremely small target size and weak backscattering responses of fishing boats, which are often indistinguishable from surrounding sea clutter. In addition, OBB regression for small targets is highly sensitive to minor localization errors, where slight deviations in angle or boundary estimation can cause a substantial drop in IoU. These factors make small-vessel detection a persistent challenge in SAR imagery, even with improved feature aggregation and task decoupling.

Regarding target integrity, YOSDet shows clear advantages for vessels with extreme scale variations. Baseline methods frequently misidentify intense local scattering on large hulls as multiple small targets, a phenomenon primarily attributed to the scale imbalance in training data. Several missed detections can still be observed in nearshore scenarios. Qualitative inspection indicates that these failures are mainly caused by extremely low signal-to-noise ratios or close proximity to high-reflectivity harbor structures, where dominant land scattering suppresses ship-specific features. This suggests that nearshore SAR ship detection remains highly sensitive to clutter intensity and local scattering complexity. Future work could explore utilizing generative adversarial networks (GANs) to augment feature representations for small targets or incorporating explicit scale supervision to mitigate such scale-dependent misidentifications. Beyond single-frame detection, the oriented geometry and high inference speed of YOSDet offer significant practical potential for temporal modeling. The orientation can serve as a critical state variable in tracking methods to refine motion prediction, while the computational efficiency provides a sufficient margin for real-time multi-frame association in continuous maritime surveillance.

6. Conclusions

In this work, we propose YOSDet, a YOLO-based oriented ship detection framework tailored for SAR imagery, addressing challenges such as background interference, speckle noise, and misalignment between classification and localization. By integrating the DAM, OGDH, and LQE, YOSDet achieves superior feature representation, decoupled optimization of classification and regression, and consistent confidence calibration. Extensive evaluations on SSDD+, HRSID, and SRSDD-v1.0 benchmarks substantiate the model's efficacy in both offshore and inshore scenes, particularly for targets with extreme scale. With only 2.15 M parameters, YOSDet balances high detection accuracy with computational efficiency, making it applicable for real-time deployment. By providing reliable OBB and consistent confidence, YOSDet serves as a foundation for downstream maritime tasks such as ship tracking. Future work will prioritize augmenting small target representation and improving scale generalization to further boost detection performance across diverse SAR scenes. Overall, YOSDet provides a practical and effective framework for oriented ship detection in complex maritime environments.

Author Contributions: Conceptualization, C.Y. and Y.S.; methodology, C.Y., Y.S. and O.-S.S.; software, C.Y.; validation, C.Y., Y.S. and O.-S.S.; formal analysis, C.Y.; investigation, C.Y.; resources, Y.S. and O.-S.S.; writing—original draft preparation, C.Y., Y.S. and O.-S.S.; writing—review and editing, C.Y., Y.S. and O.-S.S.; visualization, C.Y., Y.S. and O.-S.S.; supervision, Y.S. and O.-S.S.; funding acquisition, Y.S. and O.-S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded in part by the NRF (National Research Foundation) of Korea grant funded by the Korean government (MSIT) under Grant RS-2025-02214082 and by the MSIT, Korea, under the ITRC (Information Technology Research Center) support program (IITP-2026-RS-2023-00258639) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

1. Campbell, J.B. *Introduction to Remote Sensing*, 4th ed.; Guilford Press: New York, NY, USA, 2007.
2. Toth, C.; Józków, G. Remote sensing platforms and sensors: A survey. *ISPRS J. Photogramm. Remote Sens.* **2016**, *115*, 22–36.
3. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–43. [\[CrossRef\]](#)
4. Crisp, D.J. *The State-of-the-Art in Ship Detection in Synthetic Aperture Radar Imagery*; Department of Defence: Canberra, Australia, 2004; p. 115.
5. Li, J.; Xu, C.; Su, H.; Gao, L.; Wang, T. Deep learning for SAR ship detection: Past, present and future. *Remote Sens.* **2022**, *14*, 2712. [\[CrossRef\]](#)
6. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
7. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
8. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully convolutional one-stage object detection. *arXiv* **2019**, arXiv:1904.01355. [\[CrossRef\]](#)
9. Yaseen, M. What is YOLOv8: An in-depth exploration of the internal features of the next-generation object detector. *arXiv* **2024**, arXiv:2408.15857.
10. Khanam, R.; Hussain, M. YOLOv11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725. [\[CrossRef\]](#)
11. Lei, M.; Li, S.; Wu, Y.; Hu, H.; Zhou, Y.; Zheng, X.; Ding, G.; Du, S.; Wu, Z.; Gao, Y. YOLOv13: Real-time object detection with hypergraph-enhanced adaptive visual perception. *arXiv* **2025**, arXiv:2506.17733.
12. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [\[CrossRef\]](#)
13. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 213–229.
14. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable DETR: Deformable transformers for end-to-end object detection. *arXiv* **2020**, arXiv:2010.04159.
15. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. DETRs beat YOLOs on real-time object detection. *arXiv* **2023**, arXiv:2304.08069.
16. Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. SAR ship detection dataset (SSDD): Official release and comprehensive data analysis. *Remote Sens.* **2021**, *13*, 3690. [\[CrossRef\]](#)
17. Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; Shi, J. HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation. *IEEE Access* **2020**, *8*, 120234–120254. [\[CrossRef\]](#)
18. Lei, S.; Lu, D.; Qiu, X.; Ding, C. SRSDD-v1.0: A high-resolution SAR rotation ship detection dataset. *Remote Sens.* **2021**, *13*, 5104. [\[CrossRef\]](#)
19. Li, X.; Wang, W.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized focal loss v2: Learning reliable localization quality estimation for dense object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 11632–11641.
20. 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 (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
21. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
22. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
23. Feng, C.; Zhong, Y.; Gao, Y.; Scott, M.R.; Huang, W. TOOD: Task-aligned one-stage object detection. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 3490–3499.

24. Dai, X.; Chen, Y.; Xiao, B.; Chen, D.; Liu, M.; Yuan, L.; Zhang, L. Dynamic head: Unifying object detection heads with attentions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 7373–7382.
25. 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.
26. Ding, J.; Xue, N.; Long, Y.; Xia, G.; Lu, Q. Learning RoI transformer for oriented object detection in aerial images. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–17 June 2019; pp. 2844–2853.
27. Xie, X.; Cheng, G.; Wang, J.; Yao, X.; Han, J. Oriented R-CNN for object detection. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 3520–3529.
28. Han, J.; Ding, J.; Xue, N.; Xia, G.S. ReDet: A rotation-equivariant detector for aerial object detection. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 2786–2795.
29. Yang, X.; Yan, J.; Feng, Z.; He, T. R³Det: Refined single-stage detector with feature refinement for rotating object. *arXiv* **2019**, arXiv:1908.05612. [[CrossRef](#)]
30. Han, J.; Ding, J.; Li, J.; Xia, G.S. Align deep features for oriented object detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5602511. [[CrossRef](#)]
31. Yu, C.; Shin, Y. SAR ship detection based on improved YOLOv5 and BiFPN. *ICT Express* **2024**, *10*, 28–33. [[CrossRef](#)]
32. Yu, C.; Shin, Y. An efficient YOLO for ship detection in SAR images via channel shuffled reparameterized convolution blocks and dynamic head. *ICT Express* **2024**, *10*, 673–679. [[CrossRef](#)]
33. Zhang, T.; Zhang, X.; Ke, X. Quad-FPN: A novel quad feature pyramid network for SAR ship detection. *Remote Sens.* **2021**, *13*, 2771. [[CrossRef](#)]
34. Yu, C.; Shin, Y. SMEP-DETR: Transformer-based ship detection for SAR imagery with multi-edge enhancement and parallel dilated convolutions. *Remote Sens.* **2025**, *17*, 953. [[CrossRef](#)]
35. Sun, Z.; Leng, X.; Lei, Y.; Xiong, B.; Ji, K.; Kuang, G. BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images. *Remote Sens.* **2021**, *13*, 4209. [[CrossRef](#)]
36. Shao, Z.; Zhang, X.; Zhang, T.; Xu, X.; Zeng, T. RBFA-Net: A rotated balanced feature-aligned network for rotated SAR ship detection and classification. *Remote Sens.* **2022**, *14*, 3345. [[CrossRef](#)]
37. Yue, T.; Zhang, Y.; Wang, J.; Xu, Y.; Liu, P. A weak supervision learning paradigm for oriented ship detection in SAR image. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5207812. [[CrossRef](#)]
38. Sun, Z.; Leng, X.; Zhang, X.; Zhou, Z.; Xiong, B.; Ji, K.; Kuang, G. Arbitrary-direction SAR ship detection method for multiscale imbalance. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5208921.
39. Yasir, M.; Liu, S.; Pirasteh, S.; Xu, M.; Sheng, H.; Wan, J.; de Figueiredo, F.A.; Aguilar, F.J.; Li, J. YOLOShipTracker: Tracking ships in SAR images using lightweight YOLOv8. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *134*, 104137. [[CrossRef](#)]
40. Yasir, M.; Liu, S.; Xu, M.; Aguilar, F.J.; Wan, J.; Wei, S.; Pirasteh, S.; Fan, H.; Islam, Q.U. TFST: Two-frame ship tracking for SAR using YOLOv12 and feature-based matching. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *19*, 3175–3189. [[CrossRef](#)]
41. Yasir, M.; Liu, S.; Xu, M.; Sheng, H.; Aguilar, F.J.; do Lago Rocha, R.; de Figueiredo, F.A.P.; Colak, A.T.I.; Hossain, M.S. SSGNet: A single-source generalization model for ship detection in UAV imagery under challenging maritime environments. *Ocean Eng.* **2026**, *348*, 124120. [[CrossRef](#)]
42. Chen, Y.; Dai, X.; Liu, M.; Chen, D.; Yuan, L.; Liu, Z. Dynamic convolution: Attention over convolution kernels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11030–11039.
43. Zhu, X.; Hu, H.; Lin, S.; Dai, J. Deformable ConvNets v2: More deformable, better results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9308–9316.
44. Llerena, J.M.; Zeni, L.F.; Kristen, L.N.; Jung, C. Gaussian bounding boxes and probabilistic intersection-over-union for object detection. *arXiv* **2021**, arXiv:2106.06072.
45. Zhou, Y.; Yang, X.; Zhang, G.; Wang, J.; Liu, Y.; Hou, L.; Jiang, X.; Liu, X.; Yan, J.; Lyu, C.; et al. Mmrotate: A rotated object detection benchmark using pytorch. In Proceedings of the 30th ACM International Conference on Multimedia, Lisbon, Portugal, 10–14 October 2022; pp. 7331–7334.
46. Xu, Y.; Fu, M.; Wang, Q.; Wang, Y.; Chen, K.; Xia, G.S.; Bai, X. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 1452–1459. [[CrossRef](#)]
47. 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 (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 9759–9768.

48. Yang, Z.; Liu, S.; Hu, H.; Wang, L.; Lin, S. RepPoints: Point set representation for object detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9656–9665.
49. Li, W.; Chen, Y.; Hu, K.; Zhu, J. Oriented RepPoints for aerial object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 1829–1838.
50. Hou, L.; Lu, K.; Xue, J.; Li, Y. Shape-adaptive selection and measurement for oriented object detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 22 February–1 March 2022; Volume 36, pp. 923–932.
51. Gu, Y.; Fang, M.; Peng, D. TIAR-SAR: An oriented SAR ship detector combining a task interaction head architecture with composite angle regression. *Remote Sens.* **2025**, *17*, 2049. [[CrossRef](#)]

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