

Article

Not peer-reviewed version

CORE-Net: A Collaborative Optimization Framework for Rotated Ship Detection in Complex SAR Scenes

[Yongqi Kang](#) and [Haiping Qu](#) *

Posted Date: 5 May 2026

doi: 10.20944/preprints202605.0111.v1

Keywords: synthetic aperture radar (SAR); rotated ship detection; feature alignment; cascade regression; sample reliability



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

CORE-Net: A Collaborative Optimization Framework for Rotated Ship Detection in Complex SAR Scenes

Yongqi Kang and Haiping Qu *

School of Computer and Artificial Intelligence, Ludong University, No. 186 Hongqi Middle Road, Yantai, China

* Correspondence: quhaiping@ldu.edu.cn

Highlights

What are the main findings?

- This study proposes CORE-Net, a collaborative optimization framework that integrates RCFP, PCR Head, OAREU, and UARS to jointly address directional mismatch, unstable angle regression, and supervision-quality variation in complex SAR rotated ship detection.
- Ablation and comparative experiments demonstrate that the proposed method yields consistent improvements in $AP_{50:95}$ while preserving high Recall and Precision, with gains being particularly pronounced in inshore scenes.
- What are the implications of the main findings?
- These findings indicate that multi-level collaborative optimization constitutes an effective paradigm for enhancing both robustness and fine-grained localization quality in rotated ship detection under complex SAR conditions.
- The proposed design principles may offer instructive insights for related tasks involving rotated object detection in the presence of noisy backgrounds, arbitrary orientations, and dense target distributions.

Abstract

Rotated ship detection in complex synthetic aperture radar (SAR) scenes remains a critical yet challenging task for maritime remote sensing applications. Existing methods are plagued by three core bottlenecks: inconsistent directional responses across multi-scale features, unstable rotation angle regression, and non-uniform supervision quality of positive samples during training, which collectively lead to elevated false alarms, missed detections, and severe localization degradation, especially under high IoU thresholds in complex inshore environments. To address these challenges, we propose CORE-Net, a collaborative optimization framework integrating three dedicated modules in the forward detection stage: a Rotation-Consistent Feature Pyramid (RCFP) to alleviate cross-scale directional mismatch, a Progressive Cascade Rotation Head (PCR Head) to improve progressive angle prediction stability, and an Orientation-Aware Regression Enhancement Unit (OAREU) to strengthen directional geometric representation in regression features, alongside an Uncertainty-Aware Sample Reliability Steering (UARS) module for training-stage optimization to softly downweight the regression contribution of positive samples with high classification confidence but low geometric consistency. Extensive experiments on three public SAR ship detection datasets (RSDD-SAR, SSDD+, and RSAR) demonstrate that the proposed method consistently improves $AP_{50:95}$ while maintaining high Recall and Precision, validating that joint optimization of feature representation, rotated regression, and sample reliability is an effective strategy to enhance both the robustness and fine-grained localization capability of rotated ship detection in complex SAR scenes.

Keywords: synthetic aperture radar (SAR); rotated ship detection; feature alignment; cascade regression; sample reliability

1. Introduction

Synthetic aperture radar (SAR) is an active microwave remote sensing instrument that delivers all-weather, day-and-night, wide-swath imaging. As a pivotal imaging sensor for maritime situational awareness, SAR has demonstrated exceptional value in marine surveillance, vessel traffic management, resource monitoring, and national defense security. Automatic ship detection from SAR imagery constitutes a fundamental sensing task for constructing ocean surface target perception systems. Nevertheless, ship detection in complex sensing scenes—especially inshore and harbor areas—continues to face severe challenges. Intrinsic speckle noise, combined with heterogeneous sea-land clutter and strong scattering from port facilities, degrades the discriminability of target signatures and compromises sensor-image quality [1]. Furthermore, ship targets frequently present arbitrary orientations, pronounced scale variations, elongated aspect ratios, and dense distributions, which together render stable feature representation and precise localization considerably more difficult for downstream signal processing and data fusion pipelines [2]. Under such challenging sensing conditions, traditional detection methods and early deep-learning models that rely on horizontal bounding boxes often produce elevated false-alarm rates, missed detections, and inaccurate localization, failing to meet the demands of high-precision remote sensor interpretation [3].

To enhance the sensing capability for ship targets in complex scenes, rotated object detection has emerged as a pivotal research direction within the SAR community. Recent years have witnessed advances in rotated-box parameterization [3,4], rotation-sensitive loss functions [5,6], noise suppression and feature enhancement [7,8], and directional modeling with regression optimization [9–11]. While these methods have improved target representation, directional awareness, and rotated-box fitting accuracy for SAR image sensors, most of them focus on optimizing isolated components of the detection pipeline—for instance, strengthening feature alignment, introducing cascade regression, or augmenting directional sensitivity. For the multi-source error coupling inherent in complex sensor-imaging conditions, such piecemeal improvements often fail to unlock the full potential of holistic sensor-data fusion and processing-chain optimization.

Further analysis reveals that the performance bottleneck in sensor-based rotated ship detection does not originate from a single weak module; rather, it is intrinsically related to the coupling of errors across input interference, feature representation, rotated regression, and training supervision. Specifically, speckle noise and sea-land clutter at the sensor-imaging stage distort target representation quality. During multi-scale feature fusion, inconsistent directional responses may arise across sensor channels, complicating subsequent rotated-box modeling. Additionally, the periodicity and boundary discontinuity of the angle parameter make one-step direct regression inherently unstable for angle-sensing tasks. Moreover, the supervisory quality of positive samples varies during training, which can perturb model convergence and limit the final localization accuracy of the AI-enabled sensor system [1]. Among these issues, a critical yet often overlooked problem is the existence of positive samples that combine high classification confidence with low geometric consistency between predicted and ground-truth boxes. Such “high-confidence, low-localization-consistency” samples introduce unstable supervisory signals into the regression branch, thereby disrupting optimization and hindering localization improvements under high Intersection-over-Union (IoU) thresholds [12]. While existing training-optimization strategies—including dynamic sample assignment [13] and optimal transport assignment [14]—primarily concentrate on positive/negative sample selection heuristics, they usually neglect the fine-grained modeling and regulation of supervisory-quality discrepancies inside the already assigned positive sample set.

To address these limitations, this study proposes a collaborative optimization framework for rotated ship detection in complex SAR sensing scenes, termed CORE-Net. The framework jointly

refines the detection process across three critical dimensions of remote sensor data processing: feature representation, regression prediction, and training regulation. The goal is to enhance both the stability and the fine-grained localization capability of rotated ship detection under complex background conditions. In the forward detection stage, the proposed architecture comprises a Rotation-Consistent Feature Pyramid (RCFP), a Progressive Cascade Rotation Head (PCR Head), and an Orientation-Aware Regression Enhancement Unit (OAREU). The RCFP performs directional consistency calibration prior to multi-scale feature fusion, mitigating directional conflicts that arise during cross-level aggregation in sensor-derived feature pyramids. The PCR Head improves the stability of rotated-parameter prediction through a coarse-to-fine, two-stage angle-modeling strategy. Concurrently, the OAREU strengthens the capacity of regression features to encode target orientation and geometric structure along the regression pathway. In the training phase, we further introduce an Uncertainty-Aware Sample Reliability Steering (UARS) module. Without altering the original sample assignment protocol, UARS leverages the observed mismatch between classification confidence and geometric consistency to softly down-weight the regression contribution of potentially unreliable positive samples. This targeted regulation reduces the disruptive influence of unstable gradients on model optimization. UARS is designed to improve the training stability and the reliability of the learned sensor model. By synergistically combining forward structural enhancement with training-stage regulation, CORE-Net is designed to more effectively counteract the performance degradation induced by multi-level error coupling in complex SAR sensor-image interpretation.

The principal contributions of this study are summarized as follows:

- We propose CORE-Net, a collaborative optimization framework for rotated ship detection in complex SAR sensor scenes. It unifies multi-scale directional consistency modeling, progressive rotated regression, and training-stage sample reliability regulation within a cohesive pipeline to jointly improve detection stability and localization accuracy in challenging sensing environments.
- We design three complementary modules: a Rotation-Consistent Feature Pyramid (RCFP) for directional calibration prior to feature fusion, a Progressive Cascade Rotation Head (PCR Head) for progressive angle prediction, and an Orientation-Aware Regression Enhancement Unit (OAREU) for augmenting directional information in regression features. These modules are tailored to enhance the geometric perception capability of the AI-enabled sensor system.
- We introduce an Uncertainty-Aware Sample Reliability Steering (UARS) module. Operating online without altering the base sample assignment, UARS identifies supervisory-quality variations among positive samples and applies a soft down-weighting exclusively to the regression branch of unreliable samples, thereby enhancing training stability and the reliability of the learned sensor model.
- Extensive experimental results on public SAR sensor-dataset benchmarks—including RSDD-SAR, SSDD+, and RSAR—demonstrate that the proposed method effectively improves rotated-box localization under high-IoU thresholds while maintaining robust detection capability. Notably, CORE-Net exhibits more pronounced advantages in challenging inshore scenarios, confirming its suitability for practical marine remote-sensing applications.

2. Related Work

Rotated ship detection in complex SAR scenes encompasses multiple interconnected research areas, including multi-scale feature representation, rotated box regression, and training sample optimization. To contextualize the proposed methodology, this section reviews relevant literature across four key perspectives: feature alignment and directional modeling in rotated object detection, cascade optimization for rotated box regression, rotated ship detection methods in SAR imagery, and sample quality-aware optimization strategies during training. This analysis underscores the critical need for multi-scale directional consistency modeling, progressive angle regression, and positive-sample reliability regulation in complex scenes.

2.1. Feature Alignment and Orientation Modeling in Rotated Object Detection

A fundamental challenge in rotated object detection lies in the geometric mismatch induced by arbitrary object orientations. Conventional convolutional features extract local responses based on fixed-shaped receptive fields, which struggle to reliably capture discriminative information distributed along an object's principal axis. To mitigate this, researchers have developed various feature alignment and directional modeling techniques, broadly categorized as explicit spatial transformations or implicit feature enhancements.

Early approaches predominantly employed explicit spatial transformations to achieve directional alignment. For instance, SCRDet [15] utilizes a sampling alignment module with deformable convolution [16] to adaptively adjust sampling locations, enabling convolutional kernels to align along the target's orientation. The RoI Transformer [17] maps horizontal region proposals into rotated proposals via a dedicated learner and performs region-level feature alignment. Oriented RepPoints [18] represents objects as adaptively learned point sets and employs directional transformation functions to flexibly model the geometric structure of arbitrarily oriented targets, transitioning from box regression to point-set representation.

In multi-scale feature pyramid structures, directional modeling becomes increasingly intricate. Deep features possess strong semantic representation but low spatial resolution, with directional perception biased toward global contours. Conversely, shallow features offer rich edge and texture details, but are more susceptible to noise and local disturbances. Direct summation or concatenation of features with divergent directional preferences often introduces conflicting information into the fused output, complicating rotated parameter modeling for subsequent detection heads. To alleviate this, attention mechanisms have been integrated post-fusion. For example, FGAA-FPN [19] introduces angle-aware multi-head attention to enhance semantic interactions among high-level features by modeling relative directional relationships. FAFNet [20] notes that direct element-wise addition in conventional FPNs can introduce misaligned features, and proposes a Feature Selection and Alignment Pyramid Network (FSA-FPN). CGA-Det [11] enhances the rotational discriminability of detection-head inputs using direction-sensitive convolution.

More recently, researchers have begun explicitly addressing cross-level directional consistency during the fusion process. FAAFusion [21] leverages the rotational equivariance of the Fourier transform in the frequency domain to calibrate high-level feature orientation to the principal direction of low-level features prior to FPN fusion, thereby achieving cross-scale directional consistency modeling. ReAFFPN [22] strengthens inter-level feature fusion via rotation-equivariant attention, mitigating bottlenecks caused by semantic discontinuities and scale differences. These studies collectively indicate that explicitly modeling directional consistency within feature pyramids is beneficial for improving rotated object detection performance.

In summary, existing feature alignment methods are predominantly applied post-fusion or are confined to directional enhancement within individual feature levels. The modeling of cross-level directional consistency prior to fusion remains underexplored. Particularly in complex SAR scenes, directional conflicts between shallow textures and deep semantics can significantly impede subsequent rotated parameter modeling. Therefore, implementing directional consistency calibration before feature fusion represents a promising avenue for further investigation.

2.2. Cascade Optimization and Angle Refinement for Rotated Box Regression

Another key challenge in rotated object detection lies in the periodicity and boundary discontinuity inherent in angle parameters. For elongated SAR ship targets with large aspect ratios in SAR images, directly regressing single-stage angle regression is prone to prediction oscillation and localization inaccuracies, thereby degrading high-precision detection. To address this, coarse-to-fine cascade regression strategies have been introduced to iteratively refine angle predictions and improve localization fidelity.

In generic object detection, Cascade R-CNN [23] established a foundational paradigm for cascade regression by iteratively refining bounding boxes with detection heads operating at progressively increasing IoU thresholds. This concept has been successfully adapted for rotated

detection. Rotated Cascade R-CNN [24] extends the cascade structure to rotated boxes, progressively calibrating proposals via a rotated cascade region proposal network. S²A-Net [25] integrates feature alignment with angle refinement to achieve multi-step rotated box refinement within a single-stage framework. Methods such as Gliding Vertex [26] and Oriented R-CNN [27] similarly embody coarse-to-fine principles via strategies of "coarse horizontal localization + independent angle refinement."

Single-stage detectors are often preferred for practical SAR image analysis due to their inference speed and deployment convenience. However, realizing lightweight progressive angle modeling within a single-stage architecture remains non-trivial. R³Det [28] represents a notable single-stage rotated refinement detector that achieves higher precision through multi-stage feature iteration. Nonetheless, interactions between stages are typically limited to simple feature re-encoding, with underutilized angle priors, and the additional multi-stage processing introduces computational overhead. In contrast, general real-time detectors like RTMDet [29] rely primarily on one-step angle prediction when adapted for rotated tasks, lacking an explicit cascade refinement mechanism.

In general, existing cascade regression methods confirm that coarse-to-fine rotated box modeling is effective for mitigating angle prediction instability. Yet, most approaches remain tied to two-stage frameworks or multi-stage serial refinement structures. Lightweight progressive angle regression integrated within a single-stage detection head warrants further investigation. Furthermore, the utilization of angle priors from preceding stages remains rudimentary, leaving room for improved cross-stage transmission of directional information.

2.3. Challenges and Methods for Rotated Ship Detection in SAR Images

SAR imagery presents unique challenges—multiplicative speckle noise, strong sea–land clutter, and dense target arrangements in complex inshore backgrounds—that render rotated ship detection more demanding than in natural image scenarios. Recent research has addressed these challenges through advances in rotated box representation and loss design, feature enhancement and noise suppression, and direction-aware modeling.

In terms of direction-aware and rotation-robust modeling, studies have enhanced directional sensitivity by refining feature extractors or detection heads. CGA-Det [11] incorporates a spatial-channel-aware mechanism into the detection head to improve directional discrimination. SARFA-Net [9] enhances rotated box regression through shape-aware label assignment and refined feature alignment. These works underscore the utility of explicit direction-sensitive modeling for SAR rotated ship detection. However, the focus largely remains on the detection head itself, with insufficient consideration of cross-scale directional conflicts within the feature pyramid.

Overall, while substantial progress has been made in SAR rotated ship detection methods, there remains a relative scarcity of attention paid to pre-fusion directional consistency among multi-scale features. Similarly, explicit enhancement of directional geometric information along the regression pathway is underexplored. For robust detection under complex inshore backgrounds, jointly addressing feature directional alignment and regression feature enhancement is crucial for advancing fine-grained localization capabilities.

2.4. Sample-Quality Awareness and Optimization During Training

Effective utilization of training samples is paramount for optimizing detection model performance. Existing research on sample quality modeling and training optimization has focused on hard sample mining, dynamic label assignment, classification–localization consistency modeling, and uncertainty estimation. However, the predominant focus has been on determining which samples should be selected as positive, with relatively limited attention paid to how supervisory quality differences should be managed after positive assignment.

In hard sample mining, OHEM [33] prioritizes high-loss samples in each training batch for gradient back-propagation, intensifying learning on challenging instances. While effective, this strategy can be aggressive and potentially introduces noise in complex scenes. Focal Loss [34] elegantly addresses class imbalance by down-weighting the loss contribution of easy-to-classify samples via a modulating factor. Nevertheless, these methods primarily define difficulty based on

classification performance, lacking specificity for the more nuanced geometric regression challenges inherent in rotated detection tasks.

In dynamic label assignment, conventional fixed-IoU threshold strategies often struggle to accommodate variations in target scale and scene complexity. ATSS [35] adaptively determines the positive/negative threshold based on the statistical distribution of candidate box IoUs. OTA [36] formulates label assignment as a global optimal transport problem and demonstrates strong adaptability in dense target scenarios. Such methods significantly enhance the quality of positive/negative partitioning. Yet, once assignment is finalized, all positive samples are typically treated uniformly during training, disregarding the heterogeneous supervisory quality within the positive set.

In classification-localization consistency modeling, studies have observed that classification confidence can be incongruent with localization accuracy and have sought to rectify this via quality modeling. IoUNet [37] predicts the IoU between the predicted box and the ground-truth box as a proxy for localization quality, which can be utilized for non-maximum suppression or loss weighting. FCOS [38] introduces a "centerness" branch to quantify the degree of center deviation. Generalized Focal Loss (GFL) [39] jointly models classification scores and localization quality. These methods, however, are primarily designed for confidence calibration during inference or weight adjustment during label assignment; the soft regulation of regression supervision for already assigned positive samples during training receives comparatively less attention.

In uncertainty estimation, Bayesian deep learning techniques estimate predictive variance using methods like Monte Carlo Dropout [40] or model ensembling. This uncertainty can inform loss weighting or reliable prediction selection. However, such methods typically incur high computational cost, and the estimated uncertainty reflects global statistical uncertainty rather than the instantaneous reliability of a specific sample's supervisory signal at a given training iteration.

In summary, existing works have advanced training processes through hard sample mining, dynamic assignment, quality calibration, and uncertainty modeling. Yet, the majority concentrate on sample selection or confidence modeling, with insufficient exploitation of supervisory quality variations within the assigned positive sample set. Notably, positive samples with high classification responses but low geometric consistency can inject instability into the regression branch during training. Consequently, finer-grained training regulation predicated on reliability differences among positive samples merits deeper investigation.

Overall, prior studies have enhanced rotated object detection performance from the standpoints of feature alignment, rotated regression, SAR scene modeling, and training optimization, but most methods still focus on improving only a single stage of the pipeline. For rotated ship detection in complex SAR scenes, challenges such as multi-scale directional mismatch, unstable angle regression, and fluctuations in positive-sample supervision quality often coexist and interact during training. Therefore, a joint optimization strategy that holistically addresses feature representation, regression prediction, and training regulation holds considerable promise for further improving detection stability and fine-grained localization accuracy in complex environments.

3. CORE-Net Method

3.1. Overall Architecture

To simultaneously enhance the stability and localization accuracy of rotated ship detection in complex SAR scenes, we propose CORE-Net, a collaborative optimization framework whose overall architecture is conceptually illustrated in Figure 1. The framework comprises two core components: a forward detection pipeline and a training-stage sample reliability regulation mechanism. The forward detection pipeline consists of a feature extraction backbone, a Rotation-Consistent Feature Pyramid (RCFP), a Progressive Cascade Rotation Head (PCR Head), and an Orientation-Aware Regression Enhancement Unit (OAREU). During training, an Uncertainty-Aware Sample Reliability Steering (UARS) module is additionally introduced to stabilize regression optimization.

An input SAR image is first processed by the backbone to extract initial multi-scale features. The RCFP then calibrates the directional responses across different feature levels, mitigating cross-level directional conflicts that would otherwise degrade subsequent fusion. The aligned multi-scale features are forwarded to the detection head, where the PCR Head progressively models target position, scale, and orientation, while the OAREU augments directional geometric cues within the regression branch, thereby strengthening both the stability and discriminability of rotated-box learning. Final predictions are decoded and post-processed to yield rotated bounding boxes.

Notably, UARS is excluded from forward inference. Instead, after positive-sample assignment, it evaluates the supervision quality of positive samples on the fly based on classification confidence and geometric consistency. Positive samples that exhibit high classification confidence yet low geometric consistency are softly down-weighted exclusively in the regression branch, while classification supervision remains unaffected.

In summary, the RCFP enforces directional consistency across multi-scale features, the PCR Head improves the stability of progressive angle prediction, the OAREU enhances directional geometric representation in regression features, and the UARS module regulates positive sample reliability during training. These four modules operate at complementary stages of the detection pipeline and collectively advance rotated ship detection in challenging SAR environments.

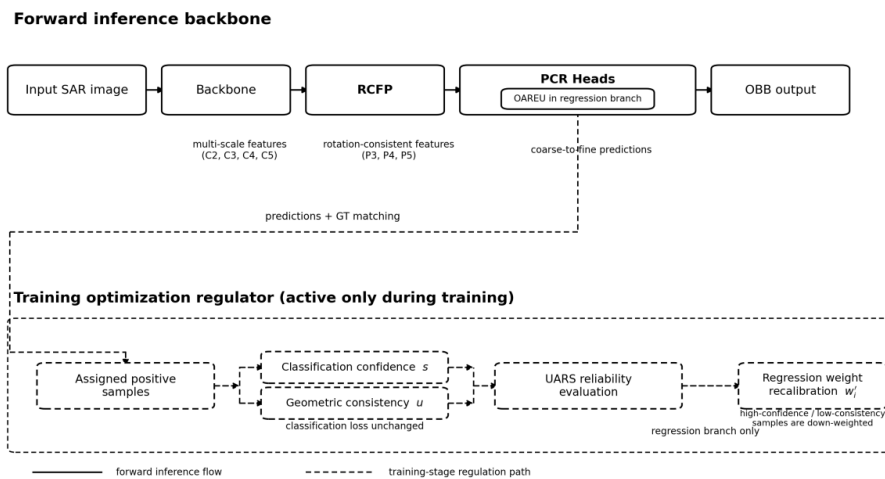


Figure 1. Overall architecture of CORE-Net.

3.2. Rotation-Consistent Feature Pyramid (RCFP)

In rotated object detection, target orientation constitutes a critical cue for inferring the geometric parameters of rotated bounding boxes, most notably the angle. However, in feature-pyramid-based multi-scale detection frameworks, the directional responses elicited by the same rotated target often vary inconsistently across feature levels. Deep features possess stronger semantic representational capacity, yet their coarse spatial resolution biases orientation perception toward global contours. Conversely, shallow features retain richer edge and texture details but are more susceptible to speckle noise, sea-land clutter, and local scattering interference. When features with disparate directional preferences are naively fused via addition or concatenation in the FPN/PAN, the resulting representations tend to embed conflicting directional information. This complicates the joint modeling of target center, scale, and angle in the subsequent detection head, often manifesting as unstable angle estimation, orientation jitter in predicted boxes, and degraded localization performance under stringent IoU thresholds. Fundamentally, this issue stems from a directional representation mismatch across feature levels prior to fusion, and it is particularly pronounced in scenarios involving elongated ships, complex near-shore backgrounds, and partial occlusion.

To mitigate this issue, we propose a Rotation-Consistent Feature Pyramid (RCFP), which calibrates the directional responses of multi-scale features prior to their input into subsequent fusion

layers and the detection head. Unlike explicit geometric rotation methods that rely on affine transformation or spatial resampling, the RCFP does not directly modify the spatial coordinates of feature maps. Instead, it performs soft directional alignment in the feature domain via three core operations: multi-directional response modeling, direction-aware weight allocation, and gated residual fusion. This design preserves the continuity of end-to-end training, while avoiding the additional computational overhead and quantization errors introduced by explicit rotational interpolation.

As illustrated in Figure 2, an identical directional consistency calibration workflow is applied to shallow, intermediate, and deep features. For an input feature map X at any level, multiple direction-sensitive branches are first employed for parallel expansion to generate a set of directional features. Each branch corresponds to a unique directional response pattern, and is responsible for capturing responses associated with the target's principal axis, edge contour, and local texture from a specific orientation. This process can be formulated as:

$$D_k = \phi_k(X), k = 1, 2, \dots, K \quad (1)$$

where $\phi_k(\cdot)$ denotes the feature mapping function of the k -th directional branch. Through this parallel directional expansion process, the original directional information entangled in the channels is explicitly decomposed into multiple candidate directional components, providing a foundational basis for subsequent direction selection and consistency modeling.

After obtaining the multi-directional responses, RCFP further introduces a lightweight direction-weight estimation module to adaptively model the importance of each directional branch. Specifically, a score is first generated for each direction based on the input feature or the multi-directional features, and then normalized to yield the corresponding directional weight α_k . This process can be written as:

$$\alpha_k = \frac{\exp(s_k)}{\sum_{j=1}^K \exp(s_j)} \quad (2)$$

where s_k represents the response score of the k -th directional branch, and α_k quantifies the relevance of that directional component to the current sample. Leveraging these weights, the multi-directional features are fused into a direction-aligned representation. Critically, this mechanism enables the network to adaptively amplify directional components that align with the target's dominant orientation while suppressing those heavily corrupted by background interference or inconsistent with the primary direction. In this manner, cross-level directional discrepancies are mitigated prior to feature fusion.

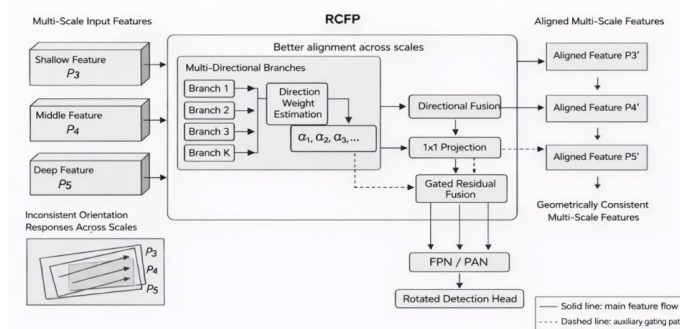


Figure 2. Structure of the Rotation-Consistent Feature Pyramid (RCFP).

Nevertheless, relying exclusively on directional aggregation risks perturbing the original feature distribution. Hence, following directional fusion, RCFP further introduces a 1×1 projection layer and

a gated residual branch. First, a 1×1 convolution projects and compresses the direction-fused feature into a compact representation. Subsequently, the projected feature Z and the original feature X are jointly fed into a gating unit to generate a gating coefficient G , after which residual reconstruction is performed. This process is formulated as:

$$G = \sigma(\gamma([Z, X])) \quad (3)$$

$$Y = G \odot Z + (1 - G) \odot X \quad (4)$$

where $\gamma(\cdot)$ denotes the gating mapping function, $\sigma(\cdot)$ is the Sigmoid function, and \odot represents element-wise multiplication. By virtue of this design, RCFP does not merely substitute the original feature with its direction-aligned counterpart. Instead, it dynamically strikes a balance between direction-enhanced information and the original response. When the directionally calibrated information proves more reliable, the gated branch increases its contribution; when the original feature is already sufficiently stable, the module retains a larger proportion of the original response. This residual fusion strategy promotes training stability while bolstering directional consistency.

Following the aforementioned processing, the input features P3, P4, and P5 are mapped into direction-calibrated aligned features P3', P4', and P5', respectively. These features are no longer mere multi-scale semantic representations; rather, they constitute detection-head inputs with substantially more consistent directional responses. The aligned multi-scale features are then passed to the subsequent FPN/PAN for further fusion and ultimately fed into the rotated detection head for bounding box prediction. Because cross-level features have already undergone preliminary directional coordination before reaching the detection head, the regression branch can devote greater capacity to fine-grained modeling of target center, scale, and angle, without additionally contending with interference arising from pre-fusion directional conflicts.

3.3. Progressive Cascade Rotation Head (PCR Head)

In conventional single-stage rotated object detectors, the target center, width, height, and rotation angle are typically regressed simultaneously by a unified detection head in a single forward pass. Although this "one-shot" prediction paradigm affords high inference efficiency, it often exhibits considerable instability when handling angle parameters that are inherently periodic and discontinuous at boundaries. For ship targets in complex SAR scenes, local features are readily corrupted by strong scattering points, speckle noise, intricate shoreline structures, and partial occlusion, rendering angle prediction overly sensitive to instantaneous feature quality. Once the initial angle estimate deviates from the true orientation, the error propagates directly into the rotated bounding box decoding process, thereby exacerbating inaccuracies in both directional estimation and localization. Consequently, replacing one-step direct angle regression with a stage-wise progressive prediction strategy constitutes an effective means of enhancing the stability of rotated box regression.

Motivated by this insight, we design a Progressive Cascade Rotation Head (PCR Head), whose overall architecture is depicted in Figure 3. In contrast to standard detection heads that output (x, y, w, h, θ) in a unified fashion, the PCR Head adopts a relatively decoupled design for the detection and angle branches. Specifically, the detection branch is responsible for category prediction and regression of the geometric parameters (x, y, w, h) , whereas the angle branch is dedicated to a two-stage cascaded prediction of the rotation angle. In implementation, the PCR Head operates on the multi-scale features F3, F4, and F5—corresponding to detection strides of 8, 16, and 32, respectively—that are fed into the detection head. Both stages share the same set of multi-scale input features, obviating the need to re-extract region features at each stage as in conventional multi-stage frameworks. This design maintains a streamlined detection pipeline while enabling progressive refinement of the angle parameter.

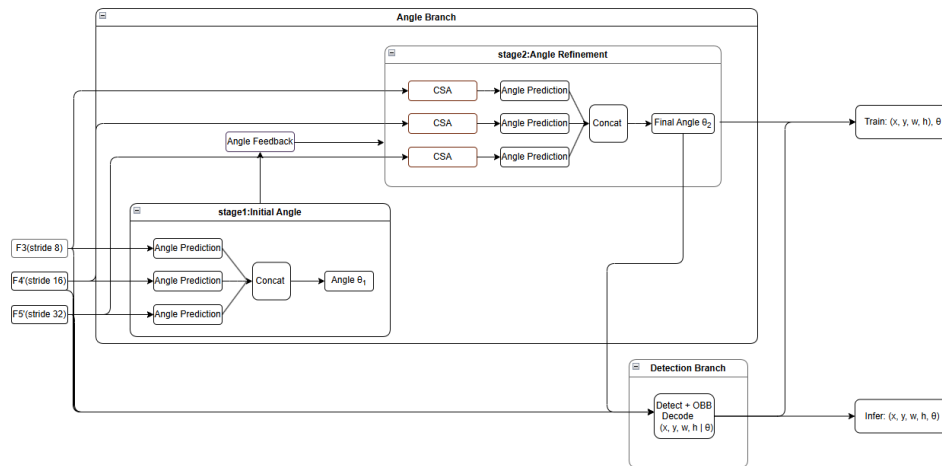


Figure 3. Cascade detection process of PCR Head.

As illustrated in Figure 3, the angle branch comprises two sequential stages. The first stage generates an initial angle estimate. For the input features F_3 , F_4 , and F_5 at different scales, each is first passed through its corresponding angle prediction sub-branch to obtain scale-specific angle responses. These responses are then flattened along the spatial dimension and concatenated to produce the initial angle prediction θ_1 of the first stage. The primary purpose of this stage is to provide a relatively stable estimate of the dominant orientation based on the current multi-scale features, thereby furnishing a directional prior for subsequent angle refinement. Concurrently, the angle responses generated at each scale in the first stage are further transformed by a lightweight mapping module to form corresponding angle embedding features, which are transmitted to the second stage via the Angle Feedback path shown in Figure 3. Thus, what the second stage receives is not the raw angle value predicted by the first stage, but rather a directional embedding representation derived from the preceding stage's angle responses.

Armed with the directional prior from the first stage, the second stage further refines the angle prediction. Rather than performing another independent angle prediction from scratch, the second stage incorporates the angle embedding features generated in the previous stage alongside the current-scale features, and accomplishes directional information transfer and feature enhancement through the Cross-Stage Attention (CSA) module, whose structure is shown in Figure 4.

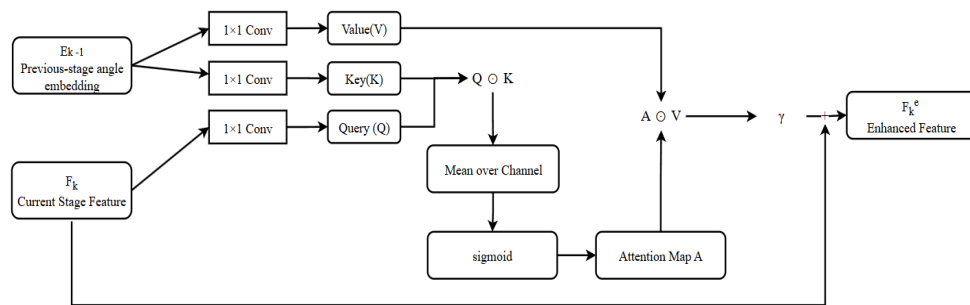


Figure 4. Structure of the cross-stage feature attention (CSA) module.

Let the current-stage feature F_k and the previous-stage angle embedding E_{k-1} be denoted as corresponding variables. They are first projected by 1×1 convolutions to obtain the query, key, and value terms:

$$Q = \phi_q(F_k), \quad K = \phi_k(E_{k-1}), \quad V = \phi_v(E_{k-1}) \quad (5)$$

where the corresponding symbols ϕ_q , ϕ_k , ϕ_v denote the respective convolution mapping functions.

Then, the element-wise interaction between Q and K is used to characterize the correlation between the current-stage feature and the directional prior from the previous stage. After aggregation along the channel dimension, a single-channel spatial response map is obtained:

$$A = \sigma(\text{Meam}_c(Q \square K)) \quad (6)$$

where the corresponding symbols \square denote element-wise multiplication, Meam_c channel-wise mean operation, and the $\sigma(\cdot)$ Sigmoid activation function, respectively.

Furthermore, the attention map A is employed to modulate the value branch V, and a learnable fusion coefficient γ is introduced to inject cross-stage information into the current-stage feature in a residual manner, yielding the enhanced stage feature:

$$F_k^e = F_k + \gamma(A \square V) \quad (7)$$

where γ is a learnable fusion coefficient used to control the strength of injecting the directional information from the previous stage into the current-stage feature. Through this design, the CSA module establishes explicit correlation modeling between the current-stage feature and the angle prior from the previous stage, thereby enhancing the utilization of directional information in the second stage.

Upon completion of cross-stage feature enhancement, the branches at different scales in the second stage output refined angle predictions, which are subsequently fused to obtain the final angle prediction θ_2 . Hence, θ_1 and θ_2 correspond, respectively, to the initial angle estimation and the refined angle prediction, forming a coarse-to-fine progressive relationship in functionality. Notably, the second stage does not explicitly learn the numerical residual of θ_1 ; instead, it outputs a more stable final angle prediction under the guidance of the directional prior supplied by the first stage. Compared with single-stage direct angle prediction, this design helps alleviate the instability of angle modeling under complex background conditions.

In addition to the angle branch, the PCR Head also retains a parallel detection branch for predicting target categories as well as center location and scale parameters. During training, the (x,y,w,h) predicted by the detection branch and the final θ_2 generated by the angle branch are jointly used to construct the supervision signal for rotated bounding boxes. During inference, the predicted geometric parameters and the final angle prediction are jointly decoded to produce the rotated detection results. In this manner, the PCR Head establishes a clear functional division: the detection branch handles position and scale modeling, while the angle branch manages coarse-to-fine directional modeling, thereby enhancing the stability of angle regression and localization accuracy without substantially increasing structural complexity.

Furthermore, a defining characteristic of the PCR Head is feature sharing. Unlike multi-stage detection frameworks in which region features are re-extracted at each stage, the two angle prediction stages in our design share the same set of multi-scale input features. Directional modulation and feature enhancement in the second stage are achieved solely through the angle embedding feedback from the previous stage and the CSA module. This design avoids the cumbersome process of repeated feature extraction and enables angle refinement with only a marginal increase in computational cost.

3.4. Orientation-Aware Regression Enhancement Unit (OAREU)

Although the RCFP mitigates cross-scale directional inconsistency at the feature pyramid level and the PCR Head reduces the sensitivity of angle prediction to one-shot estimation via progressive regression, the representational fidelity of direction-related geometric cues within the regression branch remains a decisive factor in the ultimate localization accuracy of rotated bounding boxes. For ship targets in complex SAR scenes, their geometric profiles typically manifest as elongated structures and are prone to disturbance by speckle noise, sea-land clutter, and localized strong scattering. Under such conditions, if the regression branch continues to rely predominantly on conventional convolutional features to learn bounding box parameters, several issues may ensue: insufficient directional representation, stronger entanglement between scale and angle, and amplification of local perturbations in geometric regression. These shortcomings ultimately constrain localization performance under stringent high-IOU thresholds.

To further bolster the exploitation of directional geometric information along the regression pathway, we design an Orientation-Aware Regression Enhancement Unit (OAREU) and embed it within the regression branch of the PCR Head. This module does not alter the output parameterization of the detection head. Instead, it augments the directional information of regression features at the current scale in the feature domain and incorporates auxiliary information from adjacent scales for adaptive fusion, thereby furnishing more stable and geometrically discriminative feature representations for rotated box regression. An overview of the OAREU structure is provided in Figure 5, and it primarily consists of four steps: directional enhancement, adjacent-scale auxiliary fusion, 1×1 projection, and channel recalibration.

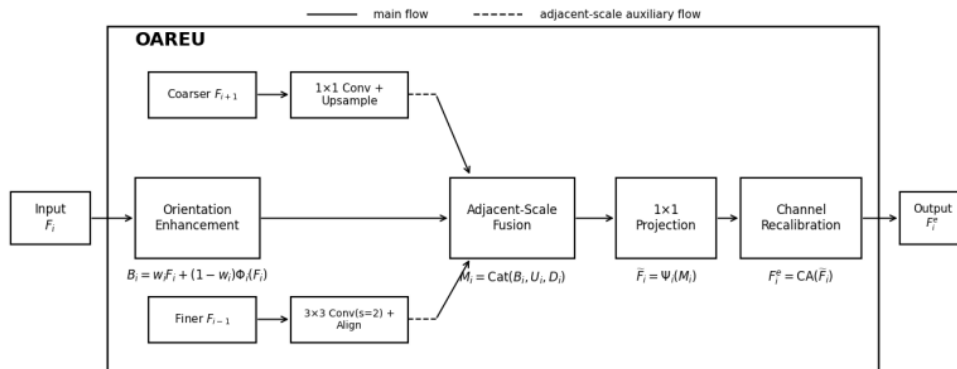


Figure 5. Structure of the OAREU module.

Let the input regression feature at the i -th detection level be denoted as F_i . First, OAREU performs directional enhancement on the current-level feature to highlight the responses related to the target principal axis, edge orientation, and local geometric structure, thereby obtaining a basic enhanced feature. It is defined as follows:

$$B_i = w_i F_i + (1 - w_i) \phi_i(F_i) \quad (8)$$

where $\phi(\cdot)$ denotes the directional enhancement transformation applied to the current-level feature, which is used to mine local structural information with stronger directional sensitivity; w_i denotes a learnable balancing coefficient used to adaptively trade off the original feature and the enhanced feature. This design avoids wholesale replacement of the original regression representation. Instead, it injects directional information that is more conducive to rotated geometric modeling while preserving the stability of the original semantic and spatial distribution.

However, relying solely on the current-scale feature for directional enhancement may still be circumscribed by the representational capacity of a single scale. Shallow features typically contain

richer edge and texture details but possess relatively weak semantic expressiveness; deep features exhibit stronger semantic abstraction yet may forfeit some fine-grained directional structure. To address this limitation, adjacent-scale auxiliary information is further introduced to complement the current-level regression representation. Concretely, the coarser-scale feature is first processed by a 1×1 convolution and an upsampling operation to generate an auxiliary feature aligned with the current scale, denoted as F_{i+1} ; the finer-scale feature is transformed by a 3×3 convolution with stride 2 and an alignment operation to obtain a downsampled auxiliary feature, denoted as F_{i-1} . Subsequently, the direction-enhanced feature of the current level and the auxiliary features from the two adjacent scales are concatenated along the channel dimension to obtain a fused representation:

$$M_i = \text{Cat}(B_i, U_i, D_i) \quad (9)$$

where $\text{Cat}(\cdot)$ denotes the channel concatenation operation. Through this step, the regression feature at the current level not only retains its own direction-sensitive information but also assimilates contextual geometric complements from adjacent scales, thereby improving its comprehensive representational capacity for elongated ship structures, scale variations, and directional continuity.

Since channel concatenation inevitably brings feature dimension expansion and a certain degree of information redundancy, a 1×1 convolution is further employed to linearly project and compress the fused result, yielding a compact intermediate representation:

$$\tilde{F}_i = \psi_i(M_i) \quad (10)$$

where $\psi_i(\cdot)$ denotes the 1×1 projection mapping. Through this process, directional and scale information from different sources is reorganized into a unified feature space, thereby reducing the redundant interference caused by direct concatenation and improving the efficiency of the subsequent regression branch in exploiting effective geometric cues.

On this basis, OAREU introduces a channel recalibration operation to adaptively reweight the projected feature and produce the final enhanced output:

$$F_i^e = \text{CA}(\tilde{F}_i) \quad (11)$$

where $\text{CA}(\cdot)$ denotes the channel recalibration function. This operation models the differential importance of distinct channels for the regression task, thereby adaptively amplifying responses more pertinent to target center localization, scale estimation, and angle prediction, while suppressing interference stemming from background clutter, noisy textures, or irrelevant scattering. The enhanced feature is ultimately fed into the subsequent regression prediction branch to support more stable learning of rotated bounding box parameters.

OAREU operates principally within the regression path. Through current-scale directional enhancement, adjacent-scale auxiliary fusion, and channel recalibration, it elevates the geometric discriminability of regression features. From an implementation standpoint, OAREU introduces only a modest number of convolution, scale alignment, and channel reweighting operations. The overall structure is lightweight and can be readily embedded into multiple levels of the regression branch without appreciably inflating the model's computational burden. Without modifying the output format of the detection head, this module augments the perception and representational capacity of regression features with respect to directional geometric information, thereby providing feature-level support for subsequent improvements in rotated box localization accuracy.

3.5. Uncertainty-Aware Sample Reliability Steering (UARS)

During the training of rotated ship detection models tailored for complex SAR scenes, the positive samples identified by the sample assigner constitute a heterogeneous set, and the supervision quality within this set often exhibits marked variation. Particularly in challenging inshore backgrounds, certain positive samples manifest a characteristic mismatch: the model yields relatively high target confidence in the classification branch, whereas the geometric consistency between the predicted rotated box and the ground-truth box remains low. For brevity, we refer to such samples as high-confidence, low-consistency (HCLC) positive samples. This phenomenon indicates that, while the model has already formed a confident detection of a target's presence at the corresponding location, the fine-grained modeling of its center, scale, and rotation angle has yet to converge to a stable state. Accordingly, the regression supervision derived from HCLC samples at the current training iteration may prove unreliable.

Such samples are highly representative of complex SAR scenes. On one hand, inshore regions are frequently afflicted by sea-land clutter, strong scattering from port infrastructure, speckle noise, and partial occlusion, conditions under which the network can still extract discriminative responses sufficient to ascertain "whether a target exists." On the other hand, ship targets themselves typically exhibit elongated profiles, arbitrary orientations, and blurred boundaries, further compounding the difficulty of accurately fitting rotated boxes in terms of position, scale, and angle. As a result, the classification and regression branches may become desynchronized for the same sample: the former may have already converged toward stability, whereas the latter may still fluctuate considerably. If such samples continue to be treated on par with high-quality positive samples during training, unstable supervision signals will be persistently injected into the regression branch, leading to oscillatory regression loss, disordered gradient directions, and ultimately constrained localization performance under high-IoU thresholds. This effect is typically more pronounced on metrics that emphasize fine localization quality, such as AP_{50:95}.

Nevertheless, directly discarding or hard-filtering such samples is equally inadvisable. High-confidence-low-consistency samples are not tantamount to erroneous samples. A substantial fraction of them remain challenging examples within the normal learning trajectory, differing only in that their geometric modeling has yet to converge at the current training stage. Indiscriminate removal might temporarily reduce noisy gradients, but it could also impair the model's adaptability to complex backgrounds, occluded targets, and ambiguous-boundary instances. Hence, such samples are better suited to a conservative and progressive regulation strategy rather than a one-shot elimination mechanism. Motivated by this observation, we propose an Uncertainty-Aware Sample Reliability Steering (UARS) module to identify and flexibly manage the regression contribution of such potentially unreliable positive samples online during training.

It is important to clarify that the term "uncertainty awareness" in UARS does not refer to explicit modeling of statistical uncertainty in the Bayesian sense, nor does it entail an additional dedicated uncertainty prediction branch. Instead, it indirectly characterizes the instability of sample supervision quality under the current training state by leveraging the discordance between classification confidence and rotated geometric consistency. In other words, UARS is not concerned with whether a sample belongs to the foreground; rather, it assesses whether the sample is suitable for participating in regression optimization with its original intensity at the current stage. Therefore, UARS neither modifies the main detector architecture nor interferes with the positive/negative sample assignment rule. Instead, it performs fine-grained reliability evaluation and training regulation within the positive sample set determined by the sample assigner. The overall workflow is depicted in Figure 6 and encompasses three primary steps: online evaluation, dynamic identification, and targeted suppression.

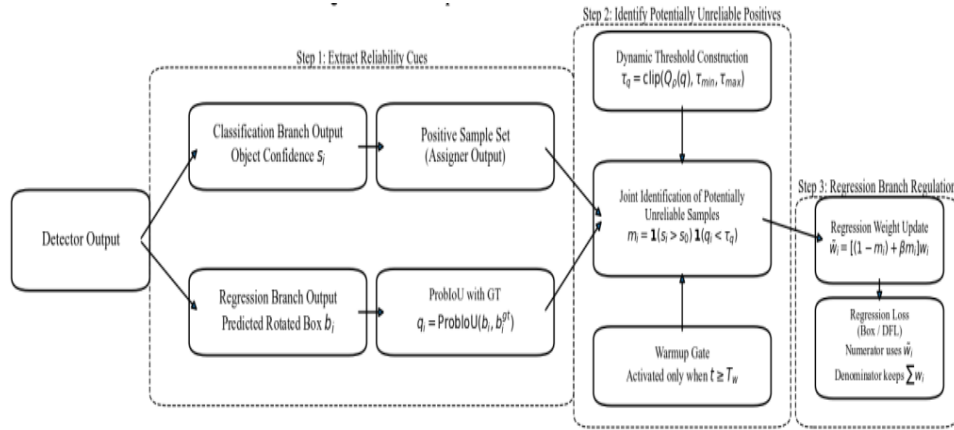


Figure 6. Overall workflow of UARS.

(1) Online Evaluation: Extracting Reliability Cues from Positive Samples

Let the positive sample set output by the sample assigner be denoted as P . For any positive sample $i \in P$, UARS extracts two complementary types of reliability cues from the classification branch and the regression branch, respectively. Specifically, the target confidence s_i is first obtained from the classification branch to quantify the model's confidence that a target exists at the corresponding location. Then, the geometric consistency is computed based on the rotated geometric similarity between the predicted rotated box b_i and the ground-truth rotated box b_i^{gt} :

$$q_i = \text{ProbIoU}(b_i^{gt}, b_i), q_i \in [0,1] \quad (12)$$

where a larger q_i indicates that the sample is better fitted in terms of position, scale, and orientation; conversely, a smaller q_i suggests that the current regression quality remains insufficient.

It warrants emphasis that UARS does not merely reiterate the original sample assignment process. The role of the sample assigner is to determine which candidate locations are deemed positive samples, whereas UARS further evaluates whether the regression supervision of these already designated positive samples is equally reliable. The former concerns foreground/background partitioning; the latter concerns quality discrimination within the positive sample set. Precisely because their functional boundaries are clearly delineated, UARS can be seamlessly integrated into the training pipeline of existing detectors as an auxiliary module without disrupting the original sample assignment rule.

(2) Dynamic Identification: Constructing the Decision Boundary for High-Confidence–Low-Consistency Samples

After obtaining s_i and q_i , the next step is to identify which positive samples should be regarded as potentially unreliable. Intuitively, such samples should satisfy two conditions simultaneously. First, the classification confidence is high, indicating that the model already strongly believes that a target exists at the corresponding location. Second, geometric consistency must be low, signifying that the fitting of the rotated box in terms of position, scale, and angle has yet to attain a level commensurate with the classification judgment. Only when both conditions are met does the sample exhibit the contradictory state wherein the classification response has stabilized while geometric regression still lags substantially behind.

In the decision strategy, UARS adopts a combination of a fixed classification threshold and a dynamic geometric threshold, rather than using fixed thresholds for both indicators. Specifically, the classification threshold s_0 is fixed, because the classification response usually becomes stable

relatively early during training, and a fixed threshold helps maintain consistency in the decision criterion. In contrast, the geometric consistency q_i fluctuates more significantly across different training stages, different batches, and different scene complexities. If a fixed geometric threshold were always used, normal difficult samples might be excessively suppressed in the early stage of training, whereas truly unstable samples might be difficult to identify effectively in the later stage. To address this issue, UARS adaptively constructs a dynamic threshold according to the distribution of geometric consistency among positive samples in the current batch:

$$\Gamma_q = \text{clip}(Q_p(\{q_i | i \in R\}), \Gamma^{\min}, \Gamma^{\max}) \quad (13)$$

where $Q_p(\cdot)$ denotes the low-quantile statistic of the geometric consistency distribution of positive samples in the current batch, and Γ^{\min} and Γ^{\max} are used to constrain the lower and upper bounds of the threshold, so as to avoid overly high or overly low threshold values under extreme distributions. In this way, Γ_q can reflect the overall geometric fitting level of the current batch while maintaining numerical stability.

Based on the above definition, the decision mask for potentially unreliable positive samples can be expressed as:

$$m_i = 1(s_i > s_0 \wedge u_i < u_0) \quad (14)$$

where $1(\cdot)$ denotes the indicator function, and u_0 denotes the threshold adaptively constructed from the consistency distribution of positive samples in the current batch. This decision logic precisely corresponds to the contradictory training scenario in which the model assigns high confidence to a sample at the classification level, yet the geometric fit between the rotated box and the ground-truth box remains deficient.

This decision logic is consistent with the schematic relationship illustrated in Figure 7. In Figure 7(a), positive samples are mapped onto a two-dimensional plane of classification confidence–geometric consistency, where the horizontal axis represents the classification confidence s_i and the vertical axis represents the geometric consistency q_i . The lower-right region partitioned by the fixed classification threshold s_0 and the dynamic geometric threshold Γ_q corresponds exactly to the UARS identification region of high-confidence but low-consistency samples. The upper-right region corresponds to reliable positive samples for which both classification judgment and geometric fitting are relatively stable. By contrast, the left region usually corresponds to ordinary positive samples or normal difficult samples whose classification confidence has not yet been sufficiently established, and therefore such samples are not suitable for focused suppression at the current stage. Figure 7(b) further illustrates the construction of the dynamic threshold Γ_q , namely, a low-quantile statistic is first extracted from the ProbIoU distribution of positive samples in the current batch, and is then constrained to a preset interval through a clip operation, thereby obtaining a geometric decision boundary that can be adaptively updated with the training state.

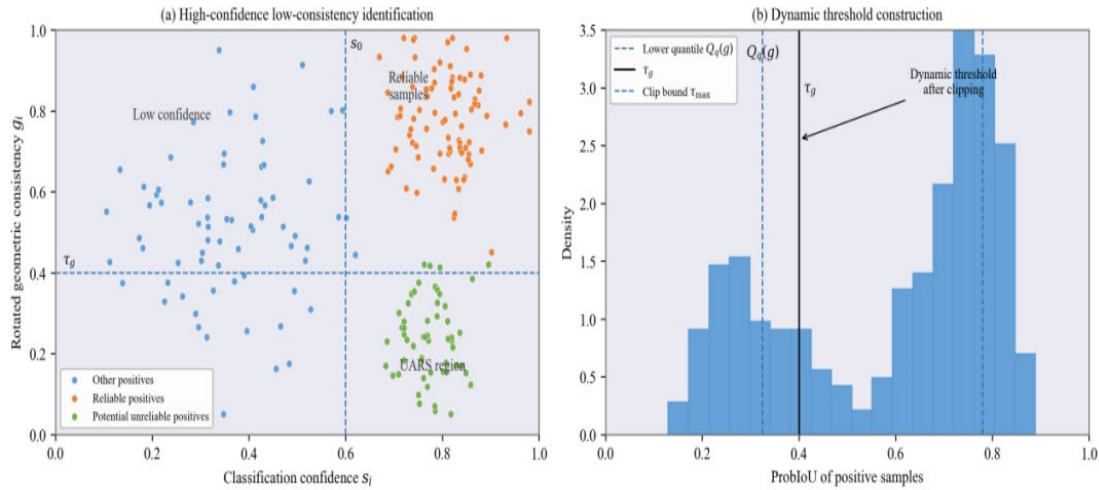


Figure 7. Illustration of positive-sample reliability identification based on high confidence and low geometric consistency.

Furthermore, UARS does not activate the above identification mechanism immediately at the onset of training. Instead, it employs a warmup gating strategy and intervenes in regression optimization only after training has entered a relatively stable phase. Let the current training epoch be denoted as t , and the warmup ending epoch be denoted as t_w . Then, UARS starts to regulate samples according to the above rule only when $t > t_w$. The reason for this design is that, in the early stage of training, the model is still in the phase of basic representation learning and preliminary geometric alignment. At that juncture, low geometric consistency is often a normative phenomenon rather than conclusive evidence of unstable supervision. Introducing suppression prematurely risks misclassifying normal difficult samples as unreliable, potentially undermining the overall convergence process.

(3) Targeted Suppression: Soft Downweighting Applied Only to the Regression Branch

Once potentially unreliable positive samples have been identified, UARS refrains from modifying the classification loss and instead applies targeted regulation solely to regression-related supervision. The underlying reason is that the core issue with such samples lies predominantly in unstable geometric learning rather than erroneous target existence judgment. Intervening in the classification branch would not only risk disturbing the already efficacious learning state of the classification pathway but would also augment the optimization coupling between classification and regression. Consequently, UARS strictly confines its intervention to the regression branch and acts only on those supervision components most prone to inducing geometric learning instability:

$$\bar{w}_i = [(1 - m_i) + \beta m_i] w_i \quad (15)$$

where β denotes the suppression coefficient, which is used to control the attenuation strength of unreliable samples in the regression loss. When $m_i = 1$, the sample is identified as a potentially unreliable sample, and its regression weight is decayed to βm_i ; when $m_i = 0$, the sample retains its original supervision strength unchanged. It can thus be seen that the essence of UARS is not to enhance high-quality samples, but rather to softly downweight high-confidence–low-consistency samples so as to reduce their disturbing effect on regression optimization.

Moreover, to ensure that the downweighting effect is not partially nullified by the loss normalization process, UARS incorporates a critical implementation detail: although the regression weight of unreliable samples is attenuated in the numerator of the loss, the normalization denominator of the regression loss is deliberately maintained as the original sum of positive sample weights $\sum_i w_i$, rather than being synchronously replaced by the downweighted $\sum_i \bar{w}_i$. If both

numerator and denominator were scaled down proportionally, the impact of downweighting on the final loss value and gradient update would be substantially diluted. By fixing the denominator and solely reducing the contribution of unreliable samples in the numerator, UARS faithfully transmits the design objective of "weakening unstable supervision" into the parameter update process. This is intuitively conveyed in Figure 8: the classification path preserves the original supervision unchanged, whereas the regression path adjusts sample weights according to the reliability decision results and utilizes them in the computation of regression-related losses such as Box Loss and DFL. Concurrently, the normalization denominator remains the original sum of positive sample supervision weights.

In summary, UARS establishes a training-level regulation mechanism that adheres to the pipeline of online evaluation \rightarrow dynamic identification \rightarrow targeted suppression. This method neither rewrites the original sample assignment rule, nor alters foreground/background partitioning, nor interferes with the classification learning pathway. Instead, it identifies those samples within the positive sample set for which the classification response has become relatively stable while geometric regression remains unstable, and softly suppresses their regression contribution during the middle and later stages of training, thereby mitigating the interference of unstable supervision on regression optimization.

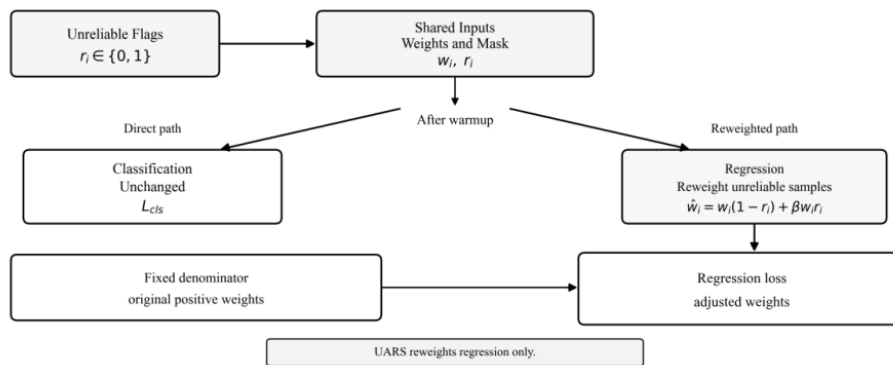


Figure 8. Reliability regulation mechanism applied only to the regression branch.

4. Experiments and Analysis

4.1. Experimental Settings

To assess the detection performance and generalization capability of the proposed method under challenging conditions, we adopted three publicly available SAR ship detection datasets as benchmarks, as listed in Table 1: RSDD-SAR [41], SSDD+ [42], and RSAR [43]. These datasets vary in scale, scene type, target distribution, and annotation format, thereby enabling a multi-faceted evaluation of model behavior under complex backgrounds, arbitrary orientations, and multi-scale variations. Notably, RSAR was originally introduced as a multi-class SAR dataset for rotated object detection. Official statistics indicate that it contains 95,842 images and 183,534 annotated instances spanning six categories: Ship, Tank, Bridge, Aircraft, Harbor, and Car. As this study concentrates on SAR ship detection, only the Ship category annotations were preserved for training and evaluation on the RSAR dataset. According to the official category statistics, the numbers of ship instances in the training, validation, and test sets are 92,950, 10,492, and 10,700, respectively, totaling 114,142 ship instances.

Table 1. Statistics of the datasets used in the experiments.

Dataset	Images	Ship Instances
RSDD-SAR	7,000	10,263
SSDD+	1,160	2,456
RSAR	95,842	114,142

Regarding data partitioning, the official splits of the public datasets were adhered to as closely as possible so as to ensure fairness and reproducibility. For RSDD-SAR and SSDD+, the publicly available test subsets were further stratified into inshore and offshore scenes in accordance with their original settings. For the RSAR dataset, the official training, validation, and test splits were adopted directly, while only the Ship annotations were retained for the single-class ship detection task. Since the official RSAR dataset lacks a demarcation between inshore and offshore scenes, only the overall test performance is reported for this dataset in the subsequent experiments. The resulting data partitioning is summarized in Table 2.

Table 2. Dataset splits used in the experiments.

Dataset	Training Images	Validation Images	Test Images	Scene Description
RSDD-SAR	5,000	-	2,000	The test set is further divided into 159 near-shore and 1,841 offshore images.
SSDD+	928	-	232	The test set is further divided into 46 near-shore and 186 offshore images.
RSAR	78,837	8,467	8,538	Official split; no near-shore/offshore division.

To facilitate a more comprehensive assessment of model performance on the SAR rotated ship detection task, Precision, Recall, AP_{50} , and $AP_{50:95}$ were employed as the principal evaluation metrics. Concretely, Precision and Recall quantify the accuracy of model predictions and the capacity to identify targets, respectively. AP_{50} captures the overall detection performance under a conventional IoU threshold, whereas $AP_{50:95}$ aggregates detection results across a spectrum of IoU thresholds and thus furnishes a more faithful reflection of the model's fine-grained localization capability for rotated bounding boxes. Given this study's emphasis on high-quality localization of rotated targets in complex scenes, $AP_{50:95}$ is regarded as the primary metric in the subsequent analysis.

All experiments were conducted in a Python-based deep learning environment, and model training and evaluation were primarily carried out within the Ultralytics framework. The versions of Ultralytics, PyTorch, and Torchvision utilized in the experiments were 8.3.203, 1.12.1, and 0.13.1, respectively, and the CUDA version was 11.3. To guarantee reproducibility, a unified input size and training regimen were maintained across all experiments. Unless otherwise specified, the training procedures were held constant across different comparative experiments, with the sole exception of the corresponding improved modules.

4.2. Module Ablation Study

To verify the independent contributions of each component and their collaborative effects, systematic ablation experiments were conducted on the three modules, namely PCR Head, RCFP, and OAREU, using YOLOv8-OB [44] as the baseline under unified training conditions. The experiments were carried out on the official offshore and inshore test subsets of the RSDD-SAR and SSDD+ datasets. Model performance was mainly evaluated using AP_{50} and $AP_{50:95}$. Except for the corresponding improved modules, all other training parameters and inference settings were kept unchanged. The detailed experimental settings are summarized in Table 3.

Table 3. Settings for the module ablation experiments of CORE-Net.

Category	Setting
Baseline Model	YOLOv8-OB
Improved Modules	PCR Head, RCFP, OAREU
Input Size	512 × 512
Optimizer	AdamW
Initial Learning Rate	0.01
Momentum	0.937
Weight Decay	0.0005
Warm-up	3 epochs
Batch Size	4
Training Epochs	50 epochs
Data Augmentation	Mosaic, random translation, scale jittering, horizontal flipping, etc.
close_mosaic	10
Inference IoU Threshold	0.7
Experimental Datasets	RSDD-SAR, SSDD+
Test Scenes	Offshore, near-shore
Core Metrics	AP ₅₀ , AP _{50:95}

The results on the RSDD-SAR dataset are presented in Table 4. Overall, all three modules produced consistent gains when introduced individually, demonstrating the effectiveness of each design. In the offshore scene, the improvement in AP_{50:95} brought by each module was relatively limited. Specifically, PCR Head, RCFP, and OAREU improved this metric from 0.7462 for the baseline to 0.7613, 0.7613, and 0.7610, respectively. By contrast, the gains in the inshore scene were more pronounced. The baseline model achieved 0.7811 and 0.5096 in terms of AP₅₀ and AP_{50:95}, respectively. After introducing PCR Head, these values increased to 0.7992 and 0.5182. With RCFP, they increased to 0.7971 and 0.5159, while with OAREU, AP_{50:95} further reached 0.5269. Moreover, multi-module combinations generally yielded better overall performance. When all three modules were enabled simultaneously, the inshore AP₅₀ and AP_{50:95} reached 0.8087 and 0.5353, respectively, representing improvements of 2.76 and 2.57 percentage points over the baseline. This indicates that the three modules exhibit good complementarity under complex backgrounds.

Table 4. Ablation results on the RSDD-SAR dataset.

PCR Head	RCFP	OAREU	Offshore Recall	Offshore Precision	Offshore AP ₅₀	Offshore AP _{50:95}	Near-shore Recall	Near-shore Precision	Near-shore AP ₅₀	Near-shore AP _{50:95}
			0.9528	0.9629	0.9855	0.7462	0.7198	0.7958	0.7811	0.5096
√			0.9573	0.9590	0.9872	0.7613	0.7215	0.7813	0.7992	0.5182
	√		0.9501	0.9689	0.9886	0.7613	0.6997	0.8143	0.7971	0.5159
		√	0.9510	0.9706	0.9882	0.7610	0.7232	0.7926	0.7968	0.5269
√	√		0.9499	0.9690	0.9875	0.7657	0.7069	0.7945	0.7879	0.5212
√		√	0.9645	0.9543	0.9877	0.7595	0.7164	0.7871	0.7965	0.5280
	√	√	0.9580	0.9635	0.9879	0.7642	0.7030	0.8140	0.7972	0.5311
√	√	√	0.9560	0.9675	0.9884	0.7621	0.7013	0.8147	0.8087	0.5353

The results on the SSDD+ dataset are summarized in Table 5, and the overall trend is largely consistent with that observed on RSDD-SAR, further demonstrating the stability of the proposed method. In the offshore scene, the effects of individual modules on AP₅₀ were relatively small, but stable gains were still observed on AP_{50:95}. In particular, OAREU and the full three-module combination achieved 0.7954 and 0.7990, respectively, both of which were higher than the baseline value of 0.7908. The improvements in the inshore scene were more evident. After jointly enabling all

three modules, AP_{50} increased from 0.8996 to 0.9188, $AP_{50:95}$ increased from 0.6023 to 0.6364, and Precision improved from 0.8419 to 0.9067. Although slight fluctuations in Recall were observed for some module combinations, the continuous improvement in Precision and $AP_{50:95}$ indicates that the main advantage of the proposed method lies in increasing the proportion of high-quality detections under complex background conditions.

Taken together, the results on both datasets show that PCR Head, RCFP, and OAREU provide stable gains from three different aspects, namely detection head modeling, multi-scale directional consistency, and rotated regression optimization. Among them, the improvements are more prominent in complex inshore scenes.

Table 5. Ablation results on the SSDD+ dataset.

PCR Head	RCFP	OAREU	Offshore Recall	Offshore Precision	Offshore AP_{50}	Offshore $AP_{50:95}$	Near-shore Recall	Near-shore Precision	Near-shore AP_{50}	Near-shore $AP_{50:95}$
			0.9840	0.9801	0.9936	0.7908	0.8358	0.8419	0.8996	0.6023
✓			0.9866	0.9857	0.9933	0.7875	0.8547	0.8352	0.9009	0.6276
	✓		0.9759	0.9849	0.9940	0.7902	0.8323	0.8721	0.9174	0.6230
		✓	0.9755	0.9892	0.9918	0.7954	0.8023	0.9193	0.8905	0.6230
✓	✓		0.9840	0.9842	0.9926	0.7851	0.8321	0.9155	0.9118	0.6225
✓		✓	0.9804	0.9920	0.9942	0.7859	0.8198	0.9235	0.9131	0.6317
	✓	✓	0.9866	0.9727	0.9934	0.7974	0.8140	0.8608	0.8987	0.6224
✓	✓	✓	0.9813	0.9905	0.9930	0.7990	0.8314	0.9067	0.9188	0.6364

To further contextualize the practical deployment cost of the proposed forward modules, the inference efficiency of varying model configurations was compared, as summarized in Table 6. The results indicate that the proposed modules incur only a moderate additional inference burden. Among these, PCR Head and OAREU exert a relatively minor impact on inference speed, whereas RCFP induces a comparatively more pronounced reduction in FPS owing to the supplementary feature-level directional calibration operations. While CORE-Net exhibits the highest inference cost among the evaluated configurations, it nonetheless maintains acceptable inference efficiency while delivering the most favorable overall detection performance.

Table 6. Inference efficiency comparison of different module configurations.

	Total Inference Time	FPS	Configuration Description
YOLOv8-OB	11.7 ms	85.47	Baseline single-stage rotated detector
PCR Head	12.4 ms	80.65	Baseline with PCR Head
RCFP	13.7 ms	72.99	Baseline augmented with feature-level RCFP
OAREU	12.2 ms	82.00	Baseline augmented with OAREU
CORE-Net	15.0 ms	66.67	Full framework with all proposed modules

On the RSDD-SAR dataset, the curves of training loss, validation loss, and $AP_{50:95}$ for the baseline model and the models equipped with different modules are shown in Figure 9(a–c). It can be observed that all models converge stably. On this basis, after introducing different modules, the fluctuations of training loss and validation loss in the middle and later stages are reduced to some extent, while the $AP_{50:95}$ curves become smoother. This indicates that the proposed method not only improves detection performance, but also contributes to enhancing training stability.

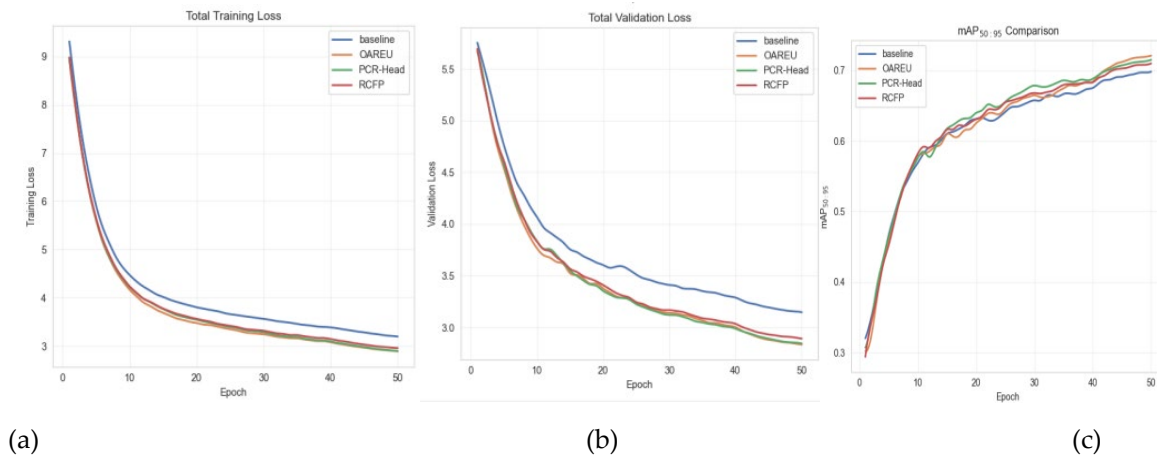


Figure 9. Comparison of training loss, validation loss, and $AP_{50:95}$ curves among the baseline model, PCR Head, RCFP, and OAREU: (a) training loss, (b) validation loss, and (c) $AP_{50:95}$.

A more granular visual analysis of the detection results across different module configurations is provided in Figure 10(a–e), wherein red boxes denote missed detections, blue boxes signify false alarms, and green boxes correspond to correctly detected targets. Taken as a whole, relative to the baseline YOLOv8-OBB, the incorporation of different modules elevates detection quality in three representative scenarios: dense distributions, complex backgrounds, and weak targets. These improvements primarily manifest as diminished missed detections, reduced false alarms, and enhanced alignment between the rotated boxes and target contours. Specific module combinations demonstrate particular strengths in directional consistency or bounding box localization, albeit with some residual localized deficiencies. In contrast, when all three modules are engaged simultaneously, the model achieves the optimal overall balance in terms of missed detection suppression, false alarm reduction, and directional consistency of rotated boxes, further attesting to the synergistic gains realized through the three modules.

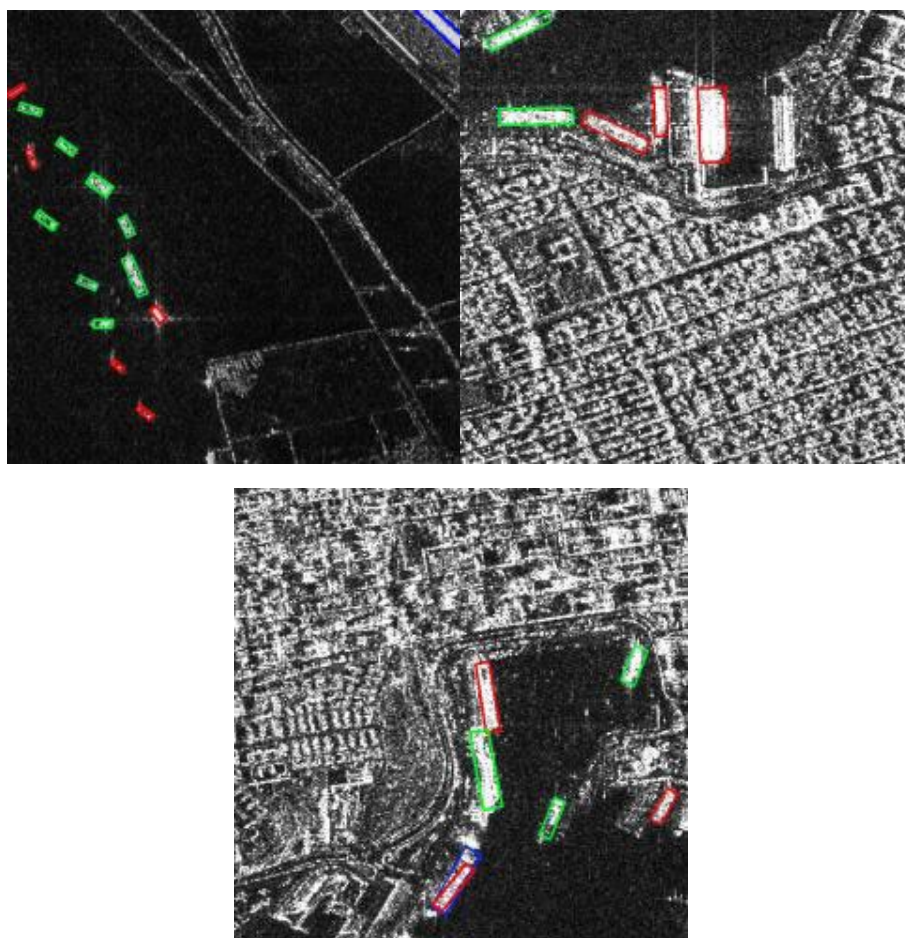


Figure 10. a. Visual comparison of detection results under different module configurations: baseline model.

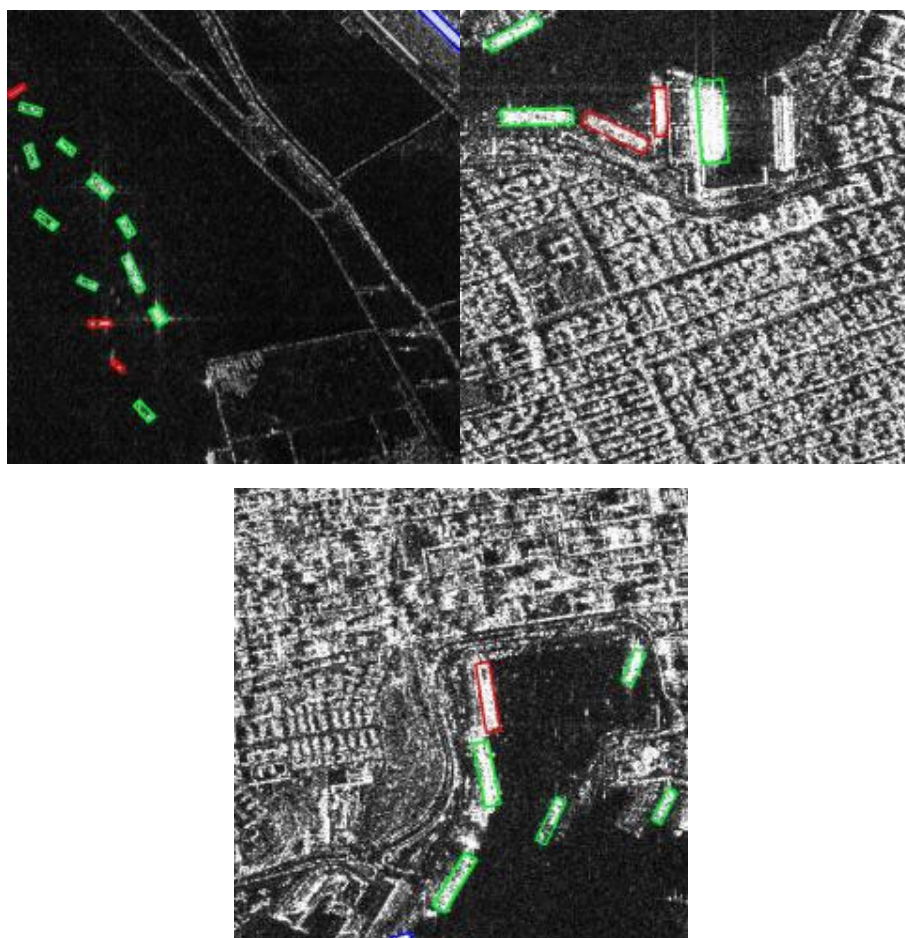


Figure 10. b. Visual comparison of detection results under different module configurations: RCFP + OAREU.

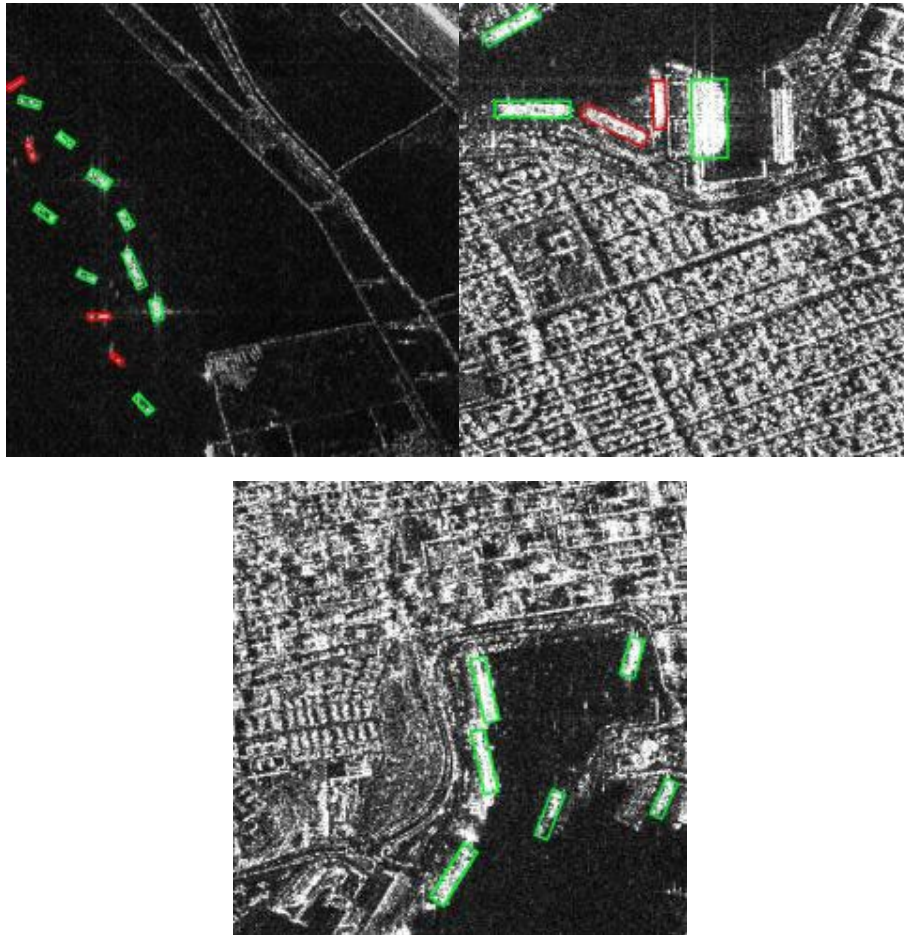


Figure 10. c. Visual comparison of detection results under different module configurations: PCR Head + RCFP.

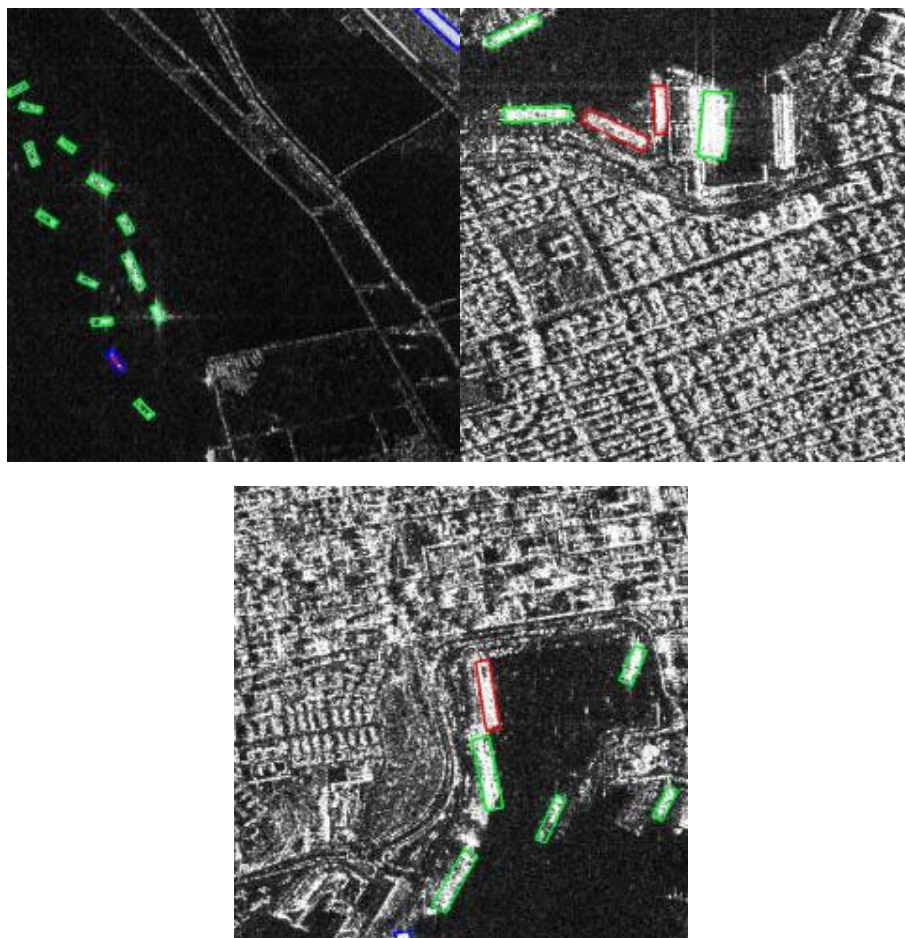


Figure 10. d. Visual comparison of detection results under different module configurations: PCR Head + OAREU.

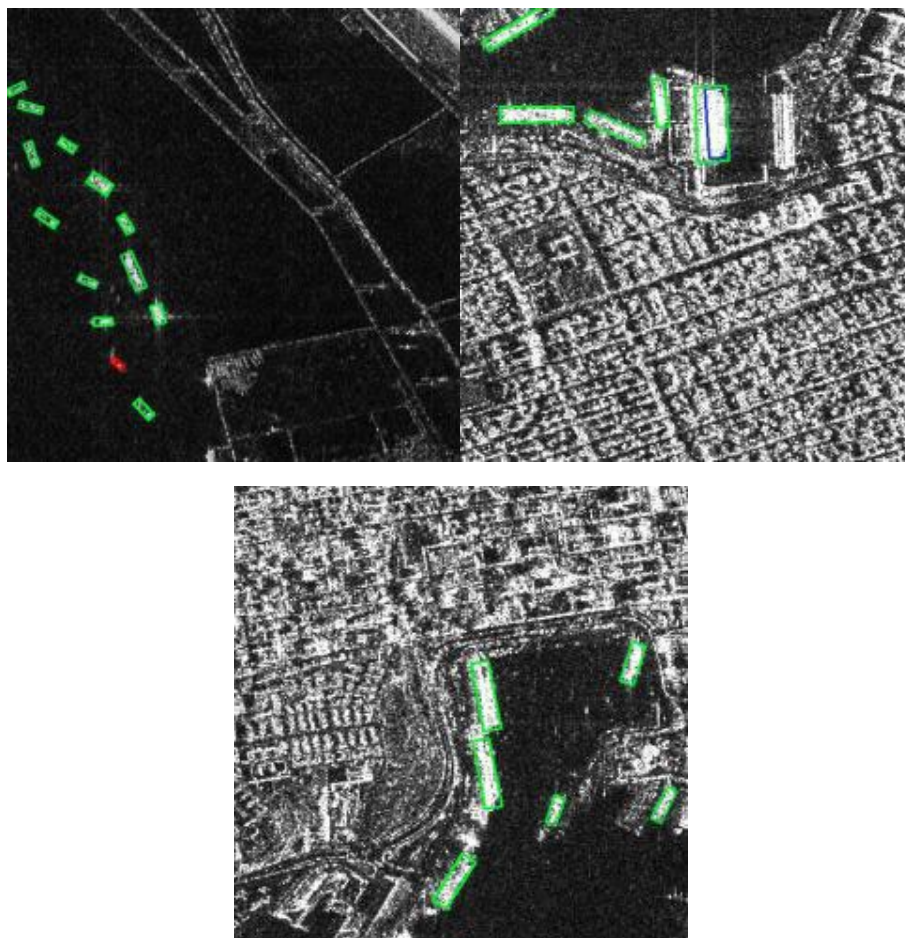


Figure 10. e. Visual comparison of detection results under different module configurations: PCR Head + RCFP + OAREU.

Through joint analysis of the quantitative ablation results and qualitative visualizations, it can be concluded that all three proposed modules deliver consistent performance improvements in both offshore and inshore scenes, with significantly more substantial gains in complex inshore environments. This performance difference can be attributed to the fact that inshore areas are typically characterized by stronger sea–land clutter, intense scattering from port facilities, and denser target distributions, all of which make conventional detection frameworks highly vulnerable to inconsistent directional responses and unstable regression. The proposed PCR Head, RCFP, and OAREU alleviate these core issues from complementary perspectives, thus delivering more pronounced gains in challenging inshore scenes.

4.3. Effectiveness Verification of UARS

To verify the independent effectiveness of UARS as a training-stage optimization strategy, as well as its generalizability across different detector architectures, UARS was integrated into both the baseline rotated detector YOLOv8-OBB and the proposed CORE-Net framework, with systematic comparative experiments conducted accordingly. The experiments primarily investigated the impact of this strategy on rotated box localization accuracy, with a specific focus on changes in $AP_{50:95}$ under high IoU thresholds. With the exception of the integration of UARS, all other network architectures, training hyperparameters, and inference settings were kept strictly fixed. Detailed experimental settings are listed in Table 7. In addition to the general training configuration, the default hyperparameter settings of UARS are also summarized in Table 7, including the fixed classification confidence threshold, warm-up duration, quantile for dynamic geometric threshold construction, clipping range of the dynamic threshold, and default suppression coefficient. Unless otherwise

specified, these hyperparameters remain fixed in all UARS-related comparative experiments. On this basis, a parameter sensitivity analysis was conducted for the core hyperparameter of UARS, the suppression coefficient β , using the inshore test subset of the RSDD-SAR dataset.

Unless otherwise stated, the results reported in Tables 8–10 are obtained using the default UARS settings listed in Table 7. Overall, UARS demonstrates excellent generalizability on both YOLOv8-OBB and the proposed CORE-Net, with performance gains predominantly concentrated on the $AP_{50:95}$ metric, underscoring its particular efficacy in enhancing fine-grained rotated-box localization under high IoU thresholds.

Table 7. Experimental settings of UARS.

Category	Setting
Baseline models	YOLOv8-OBB, CORE-Net
Comparative models	YOLOv8-OBB + UARS, CORE-Net + UARS
Input size	512 × 512
Optimizer	AdamW
Initial learning rate	0.01
Momentum	0.937
Weight decay	0.0005
Warm-up	3 epochs
Batch size	4
Training epochs	RSDD-SAR/SSDD+: 50 epochs; RSAR: 15 epochs
Data augmentation	Mosaic, random translation, scale scaling, horizontal flipping, etc.
close_mosaic	10
Inference IoU threshold	0.7
Experimental datasets	RSDD-SAR, SSDD+, RSAR
Test scenes	RSDD-SAR and SSDD+: offshore and inshore; RSAR: overall test set
Core evaluation metrics	AP_{50} , $AP_{50:95}$
Fixed confidence threshold s_0	0.60
Warm-up ending epoch t_w	8 epochs
Quantile for dynamic threshold construction q	0.20
Clipping range of dynamic threshold Γ^{\min} and Γ^{\max}	[0.25, 0.70]
Default suppression coefficient β	0.6
Application scope of UARS	Regression branch only; classification supervision remains unchanged

On the RSDD-SAR dataset, both models achieved stable improvements after introducing UARS. For YOLOv8-OBB, the $AP_{50:95}$ in the offshore scene increased from 0.7462 to 0.7549, while that in the inshore scene increased from 0.5096 to 0.5147. For CORE-Net, the offshore $AP_{50:95}$ improved from 0.7621 to 0.7645, and the inshore $AP_{50:95}$ improved from 0.5353 to 0.5406. Meanwhile, in the inshore scene, Precision increased from 0.8147 to 0.8297, and AP_{50} increased from 0.8087 to 0.8147. These results indicate that UARS plays a positive role in improving both localization quality and prediction accuracy in complex inshore scenes.

A similar trend can be observed on the SSDD+ dataset. For YOLOv8-OBB, after introducing UARS, the offshore and inshore $AP_{50:95}$ values increased from 0.7908 to 0.7954 and from 0.6023 to 0.6110, respectively. For CORE-Net, the offshore $AP_{50:95}$ improved from 0.7990 to 0.8024, while the inshore $AP_{50:95}$ improved from 0.6364 to 0.6395. At the same time, the inshore Precision increased from 0.9067 to 0.9106. Although the inshore AP_{50} of CORE-Net changed slightly from 0.9188 to 0.9179, both

AP_{50:95} and Precision were still improved, suggesting that the main benefit of UARS remains concentrated on the improvement of fine localization quality under high IoU thresholds.

The results on the RSAR dataset further verify the generalization capability of UARS. As shown in Table 10, after introducing UARS into YOLOv8-OBB, Recall, Precision, AP₅₀, and AP_{50:95} increased from 0.8144, 0.8897, 0.8900, and 0.6080 to 0.8207, 0.8921, 0.8943, and 0.6109, respectively. When UARS was further combined with CORE-Net, these metrics increased from 0.8262, 0.8911, 0.9078, and 0.6114 to 0.8307, 0.8975, 0.9118, and 0.6215, respectively. Overall, UARS does not depend on a specific detector architecture and can serve as a training-stage optimization strategy to effectively suppress unreliable positive samples in different models.

Table 8. UARS results on the RSDD-SAR dataset.

Model	Offshore Recall	Offshore Precision	Offshore AP ₅₀	Offshore AP _{50:95}	Near-shore Recall	Near-shore Precision	Near-shore AP ₅₀	Near-shore AP _{50:95}
YOLOv8-OBB	0.9528	0.9629	0.9855	0.7462	0.7198	0.7958	0.7811	0.5096
CORE-Net	0.9560	0.9675	0.9884	0.7621	0.7013	0.8147	0.8087	0.5353
YOLOv8-OBB+UARS	0.9555	0.9736	0.9872	0.7549	0.7245	0.8016	0.7888	0.5147
CORE-Net+UARS	0.9566	0.9701	0.9895	0.7645	0.7175	0.8297	0.8147	0.5406

Table 9. UARS results on the SSDD+ dataset.

Model	Offshore Recall	Offshore Precision	Offshore AP ₅₀	Offshore AP _{50:95}	Near-shore Recall	Near-shore Precision	Near-shore AP ₅₀	Near-shore AP _{50:95}
YOLOv8-OBB	0.9840	0.9801	0.9936	0.7908	0.8358	0.8419	0.8996	0.6023
CORE-Net	0.9813	0.9905	0.9930	0.7990	0.8314	0.9067	0.9188	0.6364
YOLOv8-OBB+UARS	0.9820	0.9847	0.9931	0.7954	0.8406	0.8411	0.9073	0.6110
CORE-Net+UARS	0.9819	0.9890	0.9930	0.8024	0.8358	0.9106	0.9179	0.6395

Table 10. UARS results on the RSAR dataset.

Model	Recall	Precision	AP ₅₀	AP _{50:95}
YOLOv8-OBB	0.8144	0.8897	0.8900	0.6080
CORE-Net	0.8262	0.8911	0.9078	0.6114
YOLOv8-OBB+UARS	0.8207	0.8921	0.8943	0.6109
CORE-Net+UARS	0.8307	0.8975	0.9118	0.6215

Furthermore, Table 11 presents the parameter sensitivity analysis results of the suppression coefficient β based on the CORE-Net+UARS model on the RSDD-SAR inshore test subset. It can be observed that as β increases from 0.2 to 0.8, Recall shows a gradual downward trend, whereas Precision continuously increases. This indicates that different values of β affect the regulation strength of UARS on unreliable positive samples and further alter the trade-off between Precision and Recall. In terms of overall detection performance, when $\beta=0.6$, AP₅₀ reaches the highest value of 0.8147, while AP_{50:95} also attains a relatively high level of 0.5406. When $\beta=0.8$, AP_{50:95} further increases to 0.5427, but Recall and AP₅₀ decrease slightly. Considering detection accuracy, localization quality, and parameter stability comprehensively, $\beta=0.6$ is finally selected as the default setting in this study. This outcome indicates that UARS exhibits limited sensitivity to β , and stable gains can be realized within a plausible value range.

Table 11. Sensitivity analysis of suppression coefficient β in UARS on the near-shore subset of RSDD-SAR.

β	Recall	Precision	AP ₅₀	AP _{50:95}
0.2	0.7245	0.8179	0.8044	0.5312

0.4	0.7198	0.8251	0.8041	0.5367
0.6	0.7175	0.8297	0.8147	0.5406
0.8	0.7148	0.8487	0.8129	0.5427

As shown in Figure 11, on the RSAR dataset, the loss curves of all models decrease gradually with training epochs and stabilize in the late training stage, indicating that the overall training process of all models exhibits excellent convergence. Compared with the baseline models, both YOLOv8-OBB and CORE-Net equipped with UARS exhibit smoother loss reduction trends, with significantly reduced oscillations in the middle and late training stages. This result confirms that UARS effectively mitigates the interference of unstable supervision signals on the optimization process by suppressing the regression contribution of high-confidence, low-consistency positive samples.

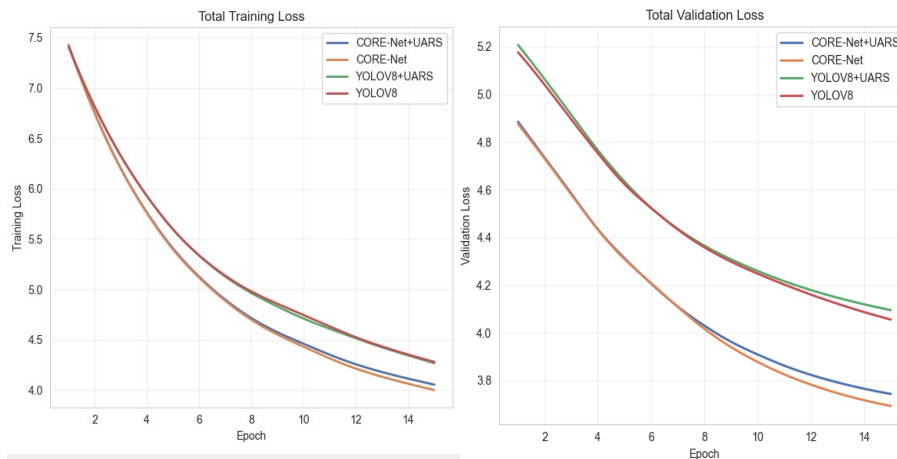


Figure 11. Loss curves of different models on the RSAR dataset.

As shown in Figure 12, after introducing UARS, the AP_{50} and $AP_{50:95}$ values of the model are usually slightly lower than those of the corresponding model without UARS in the early training stage. However, as training proceeds, the models equipped with UARS gradually surpass their counterparts in the middle and later stages, and exhibit a more stable upward trend, particularly on the $AP_{50:95}$ metric. This indicates that the main role of UARS is not to simply increase the number of detected targets, but rather to improve the rotated box localization quality under high IoU thresholds.

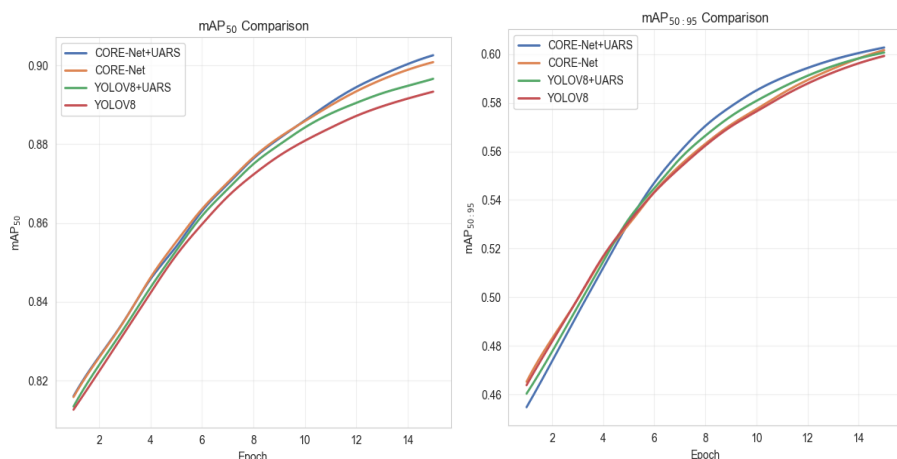


Figure 12. AP_{50} and $AP_{50:95}$ curves of different models on the RSAR dataset.

Figure 13 presents a qualitative visual comparison of detection results before and after integrating UARS. Overall, for both YOLOv8-OBB and CORE-Net, the integration of UARS reduces false detections in complex backgrounds, improves the detection stability of weak targets and targets corrupted by local interference, and produces rotated boxes with better alignment to target contours.

Figure 14 further illustrates the differences in detection results under different values of β . In general, different β values do not alter the basic detection capability of the model, but do affect false alarm suppression, missed detection control, and rotated box fitting quality to varying degrees. Among them, $\beta=0.6$ achieves the optimal balance between false alarm control, missed detection suppression, and rotated box fitting quality, which is fully consistent with the quantitative results in Table 11.

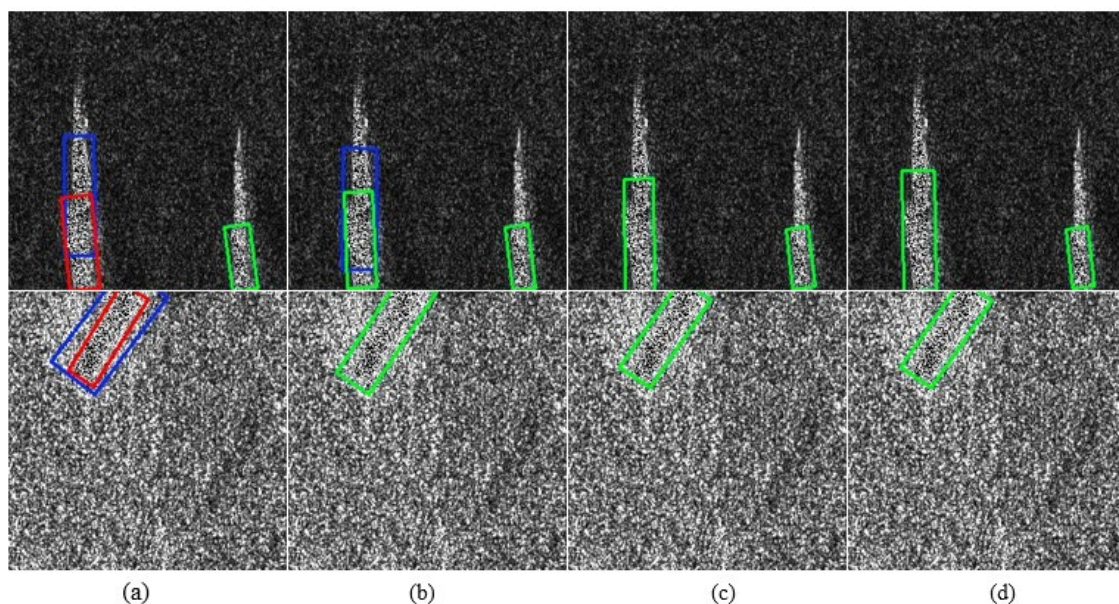


Figure 13. Visualization of detection results before and after applying UARS.(a)YOLOv8-OBB(b)YOLOv8-OBB+UARS(c)CORE-Net(d)CORE-Net+UARS.

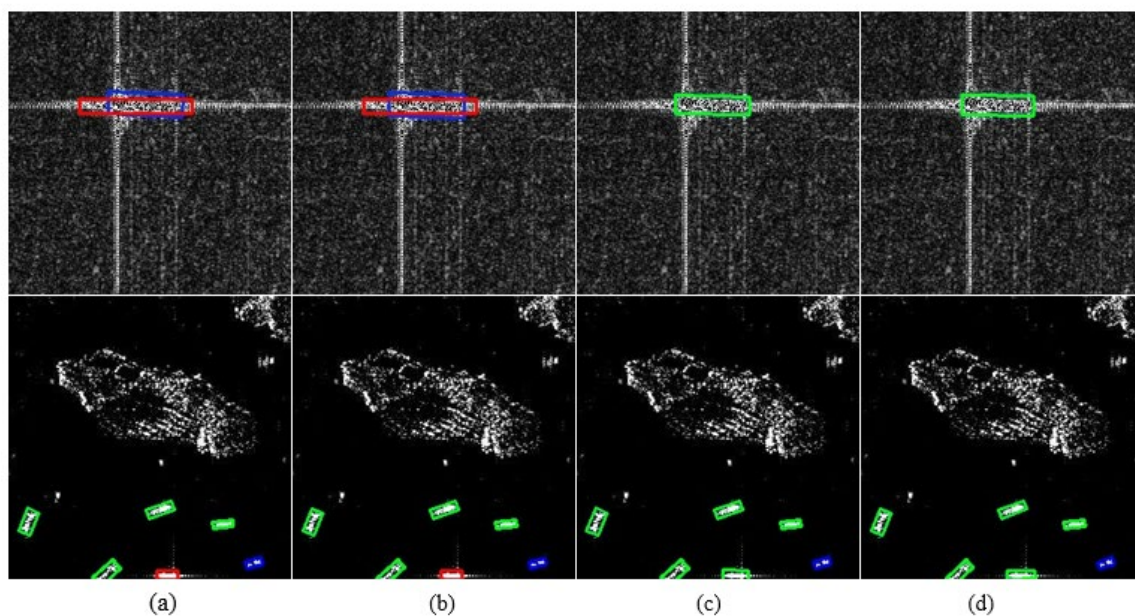


Figure 14. Visualization of detection results under different suppression coefficients β .(a) $\beta=0.2$ (b) $\beta=0.4$ (c) $\beta=0.6$ (d) $\beta=0.8$.

In summary, as a plug-and-play training-stage optimization strategy, UARS delivers consistent and stable performance gains across different detector architectures and datasets. Its core advantages are reflected in the significant improvement of $AP_{50:95}$ and the enhancement of training stability. When

combined with the proposed feature-level and detection head-level modules, UARS can further improve the fine-grained localization capability of rotated ship detection in complex SAR scenes.

4.4. Comparison with Mainstream Detectors

To further validate the competitive performance detection performance of the proposed CORE-Net+UARS framework, three mainstream rotated object detectors, namely RTMDet [29], YOLOv11-OBB [45], and YOLOv12-OBB [46], were selected as comparative baselines. Experiments were conducted on the offshore and inshore test subsets of the RSDD-SAR and SSDD+ datasets. To ensure fair and rigorous comparison, all models were evaluated under the same data partition, input resolution, and basic training pipeline. With the exception of inherent differences in model architecture, all other experimental conditions were kept strictly consistent. Quantitative results of all models are reported in Tables 12 and 13, respectively.

Table 12. Comparison with mainstream detectors on the RSDD-SAR dataset.

Model	Offshore Recall	Offshore Precision	Offshore AP ₅₀	Offshore AP _{50:95}	Near-shore Recall	Near-shore Precision	Near-shore AP ₅₀	Near-shore AP _{50:95}
RTMDet	0.9312	0.9562	0.9031	0.5225	0.5843	0.8271	0.5210	0.2842
YOLOv11-OBB	0.9464	0.9610	0.9854	0.7547	0.6795	0.7627	0.7584	0.4823
YOLOv12-OBB	0.9478	0.9576	0.9841	0.7582	0.6594	0.7949	0.7618	0.4974
CORE-Net+UARS	0.9566	0.9701	0.9895	0.7645	0.7175	0.8297	0.8147	0.5406

Table 13. Comparison with mainstream detectors on the SSDD+ dataset.

Model	Offshore Recall	Offshore Precision	Offshore AP ₅₀	Offshore AP _{50:95}	Near-shore Recall	Near-shore Precision	Near-shore AP ₅₀	Near-shore AP _{50:95}
RTMDet	0.9761	0.9732	0.9080	0.5271	0.6572	0.8432	0.6143	0.2603
YOLOv11-OBB	0.9903	0.9789	0.9942	0.7929	0.7384	0.8789	0.8606	0.5693
YOLOv12-OBB	0.9840	0.9724	0.9927	0.7878	0.7522	0.8547	0.8417	0.5528
CORE-Net+UARS	0.9819	0.9890	0.9930	0.8024	0.8358	0.9106	0.9179	0.6395

As can be seen from Tables 12 and 13, the proposed CORE-Net+UARS framework maintains high detection accuracy in the offshore scenes of both datasets, and exhibits significantly more pronounced advantages in the more challenging inshore scenes. Overall, the proposed method delivers consistent and substantial improvements in AP_{50:95}, demonstrating that its core benefits are reflected not only in target detection capability, but also in the significant enhancement of rotated box localization quality under high IoU thresholds.

In the offshore scene of the RSDD-SAR dataset, CORE-Net+UARS achieves Recall, Precision, AP₅₀, and AP_{50:95} values of 0.9566, 0.9701, 0.9895, and 0.7645, respectively, outperforming the comparison models overall. Compared with YOLOv11-OBB, AP₅₀ and AP_{50:95} are improved by 0.41 and 0.98 percentage points, respectively. Compared with YOLOv12-OBB, the corresponding improvements are 0.54 and 0.63 percentage points. These results indicate that, in offshore scenes with relatively simple backgrounds, the proposed method can still provide stable gains in high-precision rotated box fitting quality while maintaining high detection rate and prediction accuracy. In contrast, RTMDet performs markedly worse than the other OBB detectors in terms of AP₅₀ and AP_{50:95} in this

scene, suggesting that general-purpose detection frameworks still exhibit a certain adaptability gap for SAR rotated object detection tasks.

In the more challenging inshore scene of the RSDD-SAR dataset, the advantages of CORE-Net+UARS become more evident. Its Recall, Precision, AP_{50} , and $AP_{50:95}$ reach 0.7175, 0.8297, 0.8147, and 0.5406, respectively, all of which are the best among the compared models. Compared with YOLOv11-OBB, Recall, AP_{50} , and $AP_{50:95}$ are improved by 3.80, 5.63, and 5.83 percentage points, respectively. Compared with YOLOv12-OBB, Precision, AP_{50} , and $AP_{50:95}$ are improved by 3.48, 5.29, and 4.32 percentage points, respectively. In particular, the consistent advantage on $AP_{50:95}$ demonstrates that the proposed method can still maintain strong regression stability and fine-grained localization capability for rotated boxes in inshore regions affected by strong sea-land clutter, scattering from port facilities, and dense target interference. On the other hand, the performance degradation of RTMDet in the inshore scene is especially significant, with $AP_{50:95}$ reaching only 0.2842, which also indicates that complex SAR rotated ship detection imposes higher requirements on directional modeling and fine regression capability.

On the SSDD+ dataset, a trend largely consistent with that of RSDD-SAR can be observed. In the offshore scene, the Recall of CORE-Net+UARS is 0.9819, which is slightly lower than the 0.9903 of YOLOv11-OBB. However, its Precision and $AP_{50:95}$ reach 0.9890 and 0.8024, respectively, both of which are the highest among all models. Specifically, $AP_{50:95}$ is improved by 0.95 and 1.46 percentage points compared with YOLOv11-OBB and YOLOv12-OBB, respectively. Although its AP_{50} of 0.9930 is slightly lower than the 0.9942 of YOLOv11-OBB, the difference is only 0.12 percentage points, indicating that the proposed method still remains highly competitive overall. This suggests that, even in scenes with relatively simple backgrounds, the proposed method can further improve the high-precision fitting quality of rotated boxes while maintaining high detection accuracy.

For the inshore scene of SSDD+, the advantages of CORE-Net+UARS are also pronounced. Its Recall, Precision, AP_{50} , and $AP_{50:95}$ reach 0.8358, 0.9106, 0.9179, and 0.6395, respectively, all of which outperform the comparison models. Compared with YOLOv11-OBB, Recall, Precision, AP_{50} , and $AP_{50:95}$ are improved by 9.74, 3.17, 5.73, and 7.02 percentage points, respectively. Compared with YOLOv12-OBB, the corresponding improvements are 8.36, 5.59, 7.62, and 8.67 percentage points, respectively. These results indicate that, under complex inshore backgrounds, the proposed method exhibits stronger robustness than existing advanced OBB detectors. This further demonstrates that CORE-Net+UARS has better overall adaptability in terms of complex background modeling, multi-scale directional representation, and rotated regression optimization, and can more effectively alleviate false detections, missed detections, and inaccurate localization that commonly occur in inshore scenes.

Further comparative analysis across the two datasets reveals that our proposed CORE-Net+UARS achieves strong detection performance in offshore scenes, while delivering more substantial performance gains in highly complex inshore scenarios. In particular, its consistent performance superiority across the $AP_{50:95}$ metric demonstrates that the core advantage of the proposed method stems primarily from its ability to improve the fine-grained localization accuracy of rotated targets, rather than only boosting detection performance under loose, conventional IoU thresholds. More precisely, the performance gains of CORE-Net+UARS are not achieved by simply increasing the number of predicted bounding boxes; instead, the method maintains superior bounding box fitting accuracy and orientation consistency under stringent high-IoU evaluation criteria.

To further visualize the performance differences across competing methods, we conducted a qualitative comparison of detection results from all evaluated detectors in complex SAR scenes, as presented in Figure 15(a-d). The color coding follows the same convention as in Figure 10. Qualitative results show that RTMDet is prone to localization offsets, incomplete bounding box fitting, and false detections in scenarios with complex backgrounds and weak target responses. While YOLOv11-OBB and YOLOv12-OBB deliver relatively stable overall detection performance, they still exhibit non-negligible missed detections and orientation prediction deviations in dense inshore target clusters. In contrast, our proposed CORE-Net+UARS yields fewer false alarms in complex background regions,

achieves more complete detection of weak and densely distributed targets, and generates rotated bounding boxes that align more closely with the actual contours and orientations of ship targets. These qualitative observations are fully consistent with the quantitative results presented in Tables 12 and 13, further validating the superior robustness and adaptability of the proposed method in complex backgrounds, across multi-scale target variations, and for arbitrarily oriented targets.

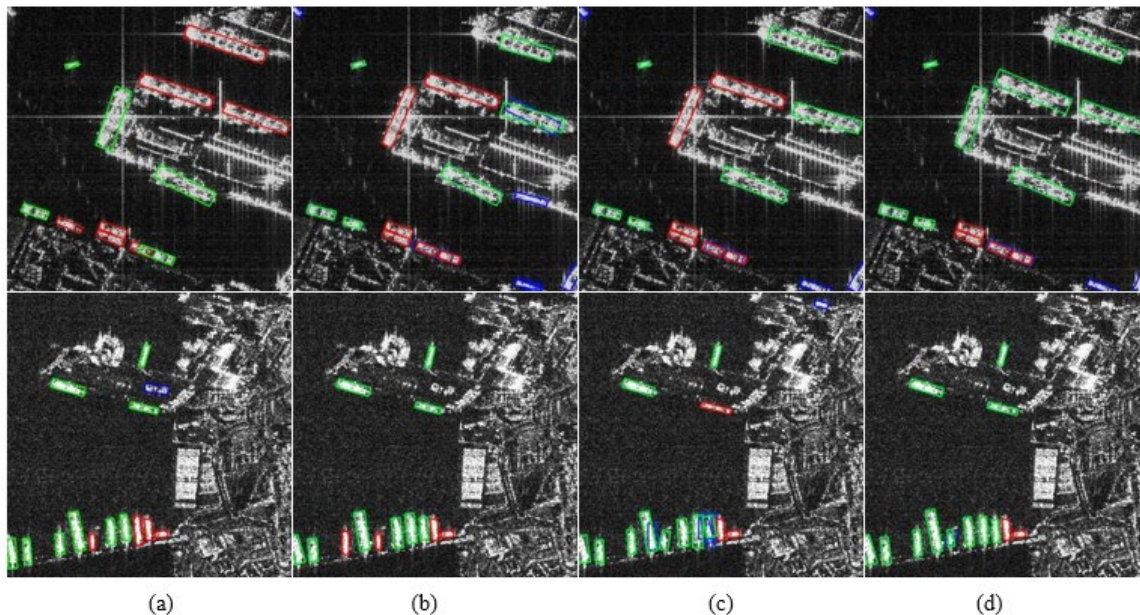


Figure 15. Qualitative comparison of detection results in complex SAR scenes.(a)RTMDet(b)YOLOv11-OB(c)YOLOv12-OB(d)CORE-Net+UARS.

5. Conclusions and Future Work

To address the localization accuracy degradation of rotated ship detection in complex SAR scenes, which is caused by multiplicative speckle noise, cross-scale directional response inconsistency, unstable rotated regression, and non-uniform supervision quality of positive samples, this study proposes a multi-level collaborative optimization framework termed CORE-Net. The proposed framework is systematically designed from three complementary and tightly coupled dimensions: feature representation, detection output modeling, and training optimization. By integrating a Rotation-Consistent Feature Pyramid (RCFP), a Progressive Cascade Rotation Head (PCR Head), an Orientation-Aware Regression Enhancement Unit (OAREU), and an Uncertainty-Aware Sample Reliability Steering (UARS) module, the framework hierarchically mitigates the core error sources of rotated ship detection in complex scenes, thereby simultaneously improving detection robustness and fine-grained localization capability.

At the feature representation level, the proposed RCFP calibrates the directional consistency of multi-scale features prior to their input into the subsequent detection pipeline, alleviating directional conflicts during cross-level feature fusion and providing directionally coordinated input representations for subsequent rotated parameter modeling. At the detection output level, we construct a rotated modeling structure composed of the PCR Head and OAREU. This design transforms one-step direct angle prediction into a progressive coarse-to-fine modeling process, while further enhancing the representation of directional geometric information along the regression path, thus improving the stability and localization accuracy of rotated box regression. At the training optimization level, the proposed UARS module addresses the non-uniform supervision quality within assigned positive samples. By exploiting the mismatch between classification confidence and geometric consistency, UARS softly suppresses the regression contribution of potentially unreliable positive samples, thereby reducing the interference of unstable supervision signals on the model convergence process.

Extensive experimental results on three public benchmark datasets—RSDD-SAR, SSDD+, and RSAR—demonstrate that the proposed method achieves state-of-the-art detection performance across both offshore and inshore scenes. Compared with the baseline model, CORE-Net yields stable performance gains in all scenarios, with more pronounced improvements observed in complex inshore environments. After further incorporating the UARS module, the model exhibits a consistent and significant improvement trend on the AP_{50:95} metric, indicating that training-stage sample reliability regulation can effectively enhance the fine-grained localization quality of rotated boxes under high IoU thresholds. Overall, through collaborative optimization across the feature representation level, detection output level, and training level, CORE-Net effectively improves the robustness and rotated localization capability of the model under complex SAR backgrounds.

Despite the promising experimental results achieved by the proposed method in complex SAR rotated ship detection, several limitations remain to be addressed in future research. First, the current study focuses exclusively on ship targets in SAR imagery, and the generalizability of the proposed framework to a wider range of SAR rotated target categories (e.g., offshore infrastructure, aircraft, and ground vehicles) requires further validation. Second, this study prioritizes the improvement of detection accuracy and localization quality, while a comprehensive quantitative analysis of model complexity, inference latency, and computational efficiency still needs to be supplemented for practical engineering deployment. Third, the cross-domain generalization capability of the model under different SAR sensors, imaging resolutions, and observation modes is also worthy of further investigation.

Future research will be conducted in the following four directions. First, we will explore a tighter integration between training-stage regulation information and forward structural optimization, to enable sample reliability cues to better guide feature alignment and rotated regression learning. Second, for more challenging scenarios—such as extremely dense target clusters, very low signal-to-clutter ratios, and severe target occlusion—we will introduce dedicated dense target separation and fine-grained feature enhancement mechanisms to further improve the model's adaptability under extreme conditions. Third, we will extend the proposed collaborative optimization strategy to other remote sensing modalities and related tasks, including rotated object detection in optical remote sensing imagery, multimodal object detection, and rotated object tracking, to verify its transferability and generality. Fourth, we will incorporate model pruning, quantization, and knowledge distillation techniques to explore feasible pathways for lightweight edge deployment while maintaining the model's performance advantages.

In summary, this work demonstrates that joint optimization across feature representation, rotated regression, and training regulation is an effective strategy for improving rotated ship detection performance in complex SAR scenes. With further improvements in cross-scene generalization, computational complexity control, and lightweight deployment, this line of research is expected to facilitate the practical application of rotated detection methods in real-world SAR intelligent interpretation tasks.

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

Funding: This research was funded by the Natural Science Foundation of Shandong Province, China (Grant No. ZR2022MF292).

Data Availability Statement: The code presented in this article is publicly available at: <https://github.com/yongqi011210/CORE-Net>. The public datasets RSDD-SAR, SSDD+, and RSAR used in this study are available from the corresponding authors of the respective datasets, with detailed information provided in the References section.

Acknowledgments: We sincerely acknowledge the contributors of the RSDD-SAR, SSDD+, and RSAR datasets for making their data publicly available, which has supported this research work.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Argenti, F.; Lapini, A.; Bianchi, T.; Alparone, L. A Tutorial on Speckle Reduction in Synthetic Aperture Radar Images. *IEEE Geoscience and Remote Sensing Magazine* 2013, 1 (3), 6–35. <https://doi.org/10.1109/MGRS.2013.2277512>
2. Jiang, Y.; Li, W.; Liu, L. R-CenterNet+: Anchor-Free Detector for Ship Detection in SAR Images. *Sensors* 2021, 21, 5693. <https://doi.org/10.3390/s21175693>
3. He, Y.; Gao, F.; Wang, J.; Hussain, A.; Yang, E.; Zhou, H. Learning Polar Encodings for Arbitrary-Oriented Ship Detection in SAR Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2021, 14, 3846–3859. <https://doi.org/10.1109/JSTARS.2021.3068530>
4. Zhang, J.; Xing, M.; Sun, G.C.; Li, N. Oriented Gaussian Function-Based Box Boundary-Aware Vectors for Oriented Ship Detection in Multiresolution SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing* 2021, 60, 5218912. <https://doi.org/10.1109/TGRS.2021.3095386>
5. Xu, Z.; Gao, R.; Huang, K.; Xu, Q. Triangle Distance IoU Loss, Attention-Weighted Feature Pyramid Network, and Rotated-SARShip Dataset for Arbitrary-Oriented SAR Ship Detection. *Remote Sensing* 2022, 14(18), 4676. <https://doi.org/10.3390/rs14184676>
6. Li, Y.; Peng, C.; Chen, Y.; Jiao, L.; Zhou, L.; Shang, R. A Deep Learning Method for Ship Detection in SAR Images with Oriented Bounding Boxes. *IEEE Transactions on Geoscience and Remote Sensing* 2021, 59 (6), 4913–4925. <https://doi.org/10.1109/TGRS.2020.3022995>
7. Zhao, M.; Zhang, X.; Kaup, A. Multitask Learning for SAR Ship Detection With Gaussian-Mask Joint Segmentation. *IEEE Transactions on Geoscience and Remote Sensing* 2023, 61, 5214516. <https://doi.org/10.1109/TGRS.2023.3304847>
8. Xu, X.W.; Zhang, X.L.; Zhang, T.W.; et al. SAR Ship Detection in Complex Scenes Based on Adaptive Anchor Assignment and IOU Supervise. *Journal of Radars* 2023, 12 (5), 1097–1111. <https://doi.org/10.12000/JR23059>
9. Li, C.; Xu, E.Z.; Samat, A.; Liu, W. SARFA-Net: Shape-Aware Label Assignment and Refined Feature Alignment for Arbitrary-Oriented Object Detection in Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2025, 18, 8865–8881. <https://doi.org/10.1109/JSTARS.2025.3532039>
10. Zhang, B.; Jiang, X.; Zhou, Y.; Liu, X.Z. A Three-Stage Cascade Rotating Regression Network for SAR Target Rotation Detection. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Kuala Lumpur, Malaysia, 2022*; pp. 1–4. <https://doi.org/10.1109/IGARSS46834.2022.9883871>
11. Zhao, C.; Fu, X.J.; Dong, J. CGA-Det: A CNN–GNN-Based Oriented SAR Ship Detector for Complex Scenes. *IEEE Geoscience and Remote Sensing Letters* 2025, 22, 3503305. <https://doi.org/10.1109/LGRS.2025.3554675>
12. Ming, Q.; Zhou, Z.; Miao, L.; Zhang, H.; Li, L. Dynamic Anchor Learning for Arbitrary-Oriented Object Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence, 2021*; Vol. 35, No. 3, pp. 2355–2363. <https://doi.org/10.1609/aaai.v35i3.16336>
13. Zhu, M.; Hu, G.; Li, S.; Zhou, H.; Wang, S.; Feng, Z. A Novel Anchor-Free Method Based on FCOS + ATSS for Ship Detection in SAR Images. *Remote Sensing* 2022, 14(9), 2034. <https://doi.org/10.3390/rs14092034>
14. Yuan, B.; Zhi, X.; Hu, J.; Zhang, W. Boosting Point Set-Based Network with Optimal Transport Optimization for Oriented Object Detection. *Remote Sensing* 2024, 16(22), 4133. <https://doi.org/10.3390/rs16224133>
15. Yang, X.; Yang, J.; Yan, J.; Zhang, Y.; Zhang, T.; Guo, Z.; Xian, X.; Fu, K. SCRDet: Towards More Robust Detection for Small, Cluttered and Rotated Objects. In *Proceedings of the IEEE/CVF International*

- Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 2019; pp. 8232–8241. <https://doi.org/10.1109/ICCV.2019.00832>
16. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017; pp. 764–773. <https://doi.org/10.1109/ICCV.2017.89>
 17. Ding, J.; Xue, N.; Long, Y.; Xia, G.S.; Lu, Q. Learning RoI Transformer for Oriented Object Detection in Aerial Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019; pp. 2849–2858. <https://doi.org/10.1109/CVPR.2019.00296>
 18. 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. <https://doi.org/10.1109/CVPR52688.2022.00187>
 19. Ma, J.; Wang, Y.; Zhang, L.; Li, Q. FGAA-FPN: Foreground-Guided Angle-Aware Feature Pyramid Network for Oriented Object Detection. arXiv 2026, arXiv:2602.10710. <https://doi.org/10.48550/arXiv.2602.10710>
 20. Liu, J.; Chen, S. A Feature-Alignment and Attention-Fusion Network for Arbitrary-Oriented Object Detection in Remote-Sensing Images. In Proceedings of the International Conference on Advances in Computer Vision Research and Applications (ACVRA 2025), Online, 4 July 2025. <https://doi.org/10.1117/12.3068184>
 21. Gu, C.; Chen, L.; Gu, L.; Fu, Y. Fourier Angle Alignment for Oriented Object Detection in Remote Sensing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), in press, 2026; pp. 1–10. <https://doi.org/10.48550/arXiv.2602.23790>
 22. Sun, C.; Xu, Y.; Wu, Z.; Wei, Z. ReAFFPN: Rotation-Equivariant Attention Feature Fusion Pyramid Networks for Aerial Object Detection. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Kuala Lumpur, Malaysia, 2022; pp. 3055–3058. <https://doi.org/10.1109/IGARSS46834.2022.9884235>
 23. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into High Quality Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018; pp. 6154–6162. <https://doi.org/10.1109/CVPR.2018.00644>
 24. Cheng, H.; Guan, N.; Wu, J.; Miao, Y.; Li, H. Learning High-Quality Bounding Box for Rotated Object Detection via Rotated Cascade Region Proposal Network. In Proceedings of the 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), 2023; pp. 308–315. <https://doi.org/10.1145/3622896.3622911>
 25. Han, J.; Ding, J.; Li, J.; Xia, G.S. Align Deep Features for Oriented Object Detection. IEEE Transactions on Geoscience and Remote Sensing 2022, 60, 1–11. <https://doi.org/10.1109/TGRS.2021.3062048>
 26. 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 Transactions on Pattern Analysis and Machine Intelligence 2021, 43 (4), 1452–1459. <https://doi.org/10.1109/TPAMI.2020.2974745>
 27. Xie, X.; Cheng, G.; Wang, J.; Yao, X.; Han, J. Oriented R-CNN for Object Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021; pp. 3520–3529. <https://doi.org/10.1109/ICCV48922.2021.00350>
 28. Yang, X.; Yan, J.; Feng, Z.; He, T. R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object. In Proceedings of the AAAI Conference on Artificial Intelligence, 2021; Vol. 35, No. 4, pp. 3163–3171. <https://doi.org/10.1609/aaai.v35i4.16426>
 29. Lyu, C.; Zhang, W.; Huang, H.; Zhou, Y.; Wang, Y.; Liu, Y.; Zhang, S.; Chen, K. RTMDet: An Empirical Study of Designing Real-Time Object Detectors. arXiv 2022, arXiv:2212.07784. <https://doi.org/10.48550/arXiv.2212.07784>
 30. Yang, X.; Yan, J.; Ming, Q.; Wang, W.; Zhang, X.; Tian, Q. Rethinking Rotated Object Detection with Gaussian Wasserstein Distance Loss. In Proceedings of the 38th International Conference on Machine Learning (ICML), Virtual, 18–24 July 2021; Vol. 139, pp. 11830–11841. <https://proceedings.mlr.press/v139/yang211.html>

31. Yang, X.; Zhou, Y.; Zhang, G.; Yang, J.; Wang, W.; Yan, J.; Zhang, X.; Tian, Q. Detecting Rotated Objects as Gaussian Distributions and Its 3-D Generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2023, 45 (4), 4335–4354. <https://doi.org/10.1109/TPAMI.2022.3197152>
32. Huang, M.; Liu, T.; Chen, Y. CViT-Net: A Convolutional and Visual Transformer Fusion Network for Small Ship Target Detection in Synthetic Aperture Radar Images. *Remote Sens.* 2023, 15, 4373. <https://doi.org/10.3390/rs15184373>
33. Shrivastava, A.; Gupta, A.; Girshick, R. Training Region-Based Object Detectors with Online Hard Example Mining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016; pp. 761–769. <https://doi.org/10.1109/CVPR.2016.89>
34. 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, 2017; pp. 2980–2988. <https://doi.org/10.1109/ICCV.2017.324>
35. 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, 2020; pp. 9759–9768. <https://doi.org/10.1109/CVPR42600.2020.00978>
36. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. OTA: Optimal Transport Assignment for Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Virtual, 2021; pp. 303–312. <https://doi.org/10.1109/CVPR46437.2021.00037>
37. Jiang, B.; Luo, R.; Mao, J.; Xiao, T.; Jiang, Y. Acquisition of Localization Confidence for Accurate Object Detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 2018; pp. 816–832. https://doi.org/10.1007/978-3-030-01264-9_48
38. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Republic of Korea, 2019; pp. 9627–9636. <https://doi.org/10.1109/ICCV.2019.00972>
39. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, Virtual, 2020; pp. 21002–21012. <https://doi.org/10.48550/arXiv.2006.04388>
40. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, New York, NY, USA, 2016; pp. 1050–1059. <https://doi.org/10.48550/arXiv.1506.02142>
41. Xu, C.A.; Su, H.; Li, J.W.; Liu, Y.; Yao, L.B.; Gao, L.; Yan, W.J.; Wang, T.Y. RSDD-SAR: Rotated Ship Detection Dataset in SAR Images. *Journal of Radars* 2022, 11 (4), 581–599. <https://doi.org/10.12000/JR22007>
42. Zhang, T.W.; Zhang, X.L.; Li, J.W.; Xu, X.W.; Wang, B.Y.; Zhan, X.; Xu, Y.Q.; Ke, X.; Zeng, T.J.; Su, H.; et al. SAR Ship Detection Dataset (SSDD): Official Release and Comprehensive Data Analysis. *Remote Sensing* 2021, 13 (18), 3690. <https://doi.org/10.3390/rs13183690>
43. Zhang, X.; Yang, X.; Li, Y.; Yang, J.; Cheng, M.-M.; Li, X. RSAR: Restricted State Angle Resolver and Rotated SAR Benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 10–17 June 2025; pp. 7416–7426. <https://doi.org/10.1109/CVPR52734.2025.00695>
44. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLO (Version 8.0.0) [Computer software]. Ultralytics, 2023. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 10 April 2026).
45. Khanam, R.; Hussain, M. YOLOv11: An Overview of the Key Architectural Enhancements. *arXiv* 2024, arXiv:2410.17725. <https://doi.org/10.48550/arXiv.2410.17725>
46. Tian, Y.; Ye, Q.; Doermann, D. YOLOv12: Attention-Centric Real-Time Object Detectors. *arXiv* 2025, arXiv:2502.12524. <https://doi.org/10.48550/arXiv.2502.12524>

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.