remote sensing

MDPI

Article

CCAI-YOLO: A High-Precision Synthetic Aperture Radar Ship Detection Model Based on YOLOv8n Algorithm

Hui Liu [1,2,*], Haoyu Dong [1], Hongyin Shi [1,2] and Fang Li [1,2]

[1] School of Intelligence Science and Technology, Beijing University of Civil Engineering and Architecture, Beijing 102616, China; 2108540023065@stu.bucea.edu.cn (H.D.); shihongyin@bucea.edu.cn (H.S.); lifang@bucea.edu.cn (F.L.)
[2] Beijing Key Laboratory of Super Intelligent Technology for Urban Architecture, Beijing University of Civil Engineering and Architecture, Beijing 102616, China
* Correspondence: liuhui@bucea.edu.cn; Tel.: +86-138-1039-1342

Highlights

What are the main findings?

- This paper proposes CCAI-YOLO, a novel SAR ship detection model based on an enhanced YOLOv8 architecture.
- CCAI-YOLO employs a synergistically optimized design to boost accuracy and robustness in noisy, complex, and multi-scale scenarios.

What is the implication of the main finding?

- Evaluations on SSDD and SAR-Ship-Dataset confirm its state-of-the-art overall performance.
- The CCAI-YOLO model effectively addresses inconsistent multi-scale fusion and target omission in complex environments.

Abstract

To tackle core challenges in detecting ship targets within synthetic aperture radar (SAR) images—including coherent speckle noise interference, complex background clutter, and multi-scale target distribution—this paper proposes a high-accuracy detection model, CCAI-YOLO. This model is based on the YOLOv8n framework, achieving systematic enhancements through the collaborative optimisation of key components: within the backbone network, the original C2f structure is replaced with the dynamic convolution module C2f-ODConv, improving the model's extraction capabilities under noisy interference; the C2f-ACmix module is integrated into the neck network, introducing a self-attention mechanism to strengthen global context information modelling, thereby better distinguishing targets from structured backgrounds; the ASFF detection head optimises multi-scale feature fusion, enhancing detection consistency across different-sized targets. Concurrently, the Inner-SIoU loss function further improves bounding box regression accuracy and accelerates convergence. Experimental results demonstrate that on the public datasets SSDD and SAR-Ship-Dataset, CCAI-YOLO achieves consistent improvements over the baseline model YOLOv8n across key metrics including F1 score, mAP50, and mAP50-95. Its overall performance surpasses current mainstream SAR ship detection methods, providing an effective solution for robust and efficient ship detection in complex scenarios.

Keywords: ship detection; synthetic aperture radar (SAR); deep learning; YOLOv8n

Academic Editor: Zenghui Zhang

Received: 24 October 2025
Revised: 23 December 2025
Accepted: 28 December 2025
Published: 1 January 2026

1. Introduction

As an active microwave imaging sensor, Synthetic Aperture Radar (SAR) possesses unique all-weather, all-time, and long-range imaging advantages, enabling its significant role in marine monitoring [1], maritime traffic management [2] and ship salvage [3]. The automatic identification and localization of ship targets from complex maritime backgrounds, known as ship target detection [4], is a central technology in SAR image interpretation with substantial research value and wide applicability.

Conventional methods for detecting ship targets in SAR imagery mainly depend on hand-crafted feature extraction algorithms and classifiers. These include Constant False Alarm Rate (CFAR) detection [5], threshold segmentation [6], adaptive detection [7], and wavelet-based detection techniques [8]. Such approaches typically identify ships by extracting their geometric, textural, and scattering characteristics. However, the fundamental limitation of such approaches lies in their reliance on the 'manual feature design-shallow classification' paradigm, which presents an inherent and irreconcilable contradiction with the highly complex physical imaging characteristics inherent in SAR imagery. This contradiction manifests particularly acutely in the following three typical challenges. Firstly, at the feature level, manually designed features are highly sensitive to coherent speckle noise prevalent in SAR imagery and struggle to adaptively capture the substantial appearance variations exhibited by targets due to changes in attitude and size. Secondly, at the model level, traditional classifiers possess limited learning capabilities, rendering them incapable of modelling the high-dimensional, non-linear feature relationships that emerge from the intertwining of ships and port facilities within complex coastal backgrounds. Consequently, false alarm rates increase significantly under conditions of severe background clutter. Finally, at the methodological level, most conventional approaches employ isolated 'detection-discrimination' workflows, failing to achieve end-to-end optimisation from image to target. This leads to cumulative errors across processing stages when handling multi-scale, densely clustered, or partially occluded targets, resulting in insufficient overall robustness. Therefore, although a series of improved algorithms represented by CFAR [9–11] have enhanced performance in specific scenarios through dynamic thresholds and multi-scale modelling, they have not overcome the inherent limitations of the 'artificial feature design' paradigm. With the growing demand for processing large-scale, highly complex SAR data, the shortcomings of traditional methods in detection accuracy, adaptability, and computational efficiency have become increasingly apparent. The core focus of current research has shifted towards leveraging deep learning approaches to construct high-level feature representations directly from data through end-to-end learning. These representations exhibit robustness to noise and strong discriminative power against background clutter, thereby systematically overcoming these long-standing challenges.

In 2012, AlexNet [12], a deep convolutional neural network (CNN)-based architecture, achieved a decisive victory in the ImageNet image recognition competition, catalyzing broad interest in deep learning methodologies. Deep learning-based target detection algorithms are generally divided into two categories depending on the use of explicit region proposals: two-stage target detection algorithms and one-stage target detection algorithms. Two-stage algorithms convert the detection task into a classification problem by first generating region proposals and then classifying the localized image regions within them. Representative models in this category include R-CNN [13], Faster R-CNN [14], Cascade R-CNN [15] and Mask R-CNN [16]. In the context of SAR ship detection, Xiao et al. [17] introduced a multi-resolution detection approach based on an improved regional convolutional neural network (R-CNN), which enhanced input image sizing, region proposal optimization, database categorization and weight balancing to boost detection accuracy in complex multi-resolution SAR scenarios. In another study, Ke et al. [18] incor-

porated deformable convolutions into Faster R-CNN, enabling the model to adaptively learn two-dimensional offsets and better represent geometric variations in ship shapes, which raised average precision. Jian et al. [19] developed an SS R-CNN framework using self-supervised learning, where feature representation networks were pre-trained on ship-free ocean imagery, leading to improved Mask R-CNN performance in remote sensing ship detection, particularly for small ships. Two-stage object detection algorithms first generate region proposals, followed by classification and refinement. This approach tends to produce numerous false alarm proposals in terrestrial areas, consuming significant computational resources in subsequent processing stages. Consequently, it struggles to meet the practical demands for rapid processing of large-scale SAR data. In contrast, one-stage target detection algorithms, also referred to as regression-based detectors, bypass explicit region proposal generation and treat detection as a unified regression problem over the entire image. Notable one-stage models include YOLO [20], SSD [21], RetinaNet [22] and FCOS [23]. For SAR ship detection, Yu et al. [24] augmented YOLOv5 with a coordinate attention mechanism and a bidirectional feature pyramid, improving both feature fusion and detection accuracy. Miao et al. [25] combined wavelet decomposition with an enhanced SSD model to strengthen the detection of coastal ships in complex SAR environments. Yang et al. [26] proposed an improved fully convolutional one-stage detector (Improved-FCOS) incorporating multi-level feature attention, feature refinement reuse, and enhanced detection heads, addressing issues such as misclassification, small object detection and anchor-related limitations in SAR imagery. One-stage object detection algorithms unify detection as a dense prediction problem, significantly enhancing efficiency. However, this approach leads to conflicting optimisation objectives in complex terrestrial and open-sea environments, making it challenging to simultaneously maintain high detection rates and low false alarm rates. Although both categories of deep learning approaches have propelled progress in the field, existing methods have failed to systematically model and resolve core challenges inherent in SAR imagery, such as speckle noise, multi-scale targets, and the extreme heterogeneity of sea–land backgrounds. Consequently, developing specialised detection architectures capable of deeply integrating SAR imaging mechanisms to achieve a superior balance between accuracy, speed, and scene adaptability has become an urgent research priority requiring breakthroughs.

To address the core challenges encountered in SAR ship detection, this paper proposes a novel detection model, CCAI-YOLO. Building upon YOLOv8n as its baseline, this model avoids a simple stacking of components. Instead, through a series of targeted designs, it constructs a systematic solution that collaboratively tackles the unique difficulties inherent in SAR imagery. The principal contributions of this study are summarised as follows:

1. A novel ship detection model named CCAI-YOLO is proposed, demonstrating superior multi-scale ship detection capabilities in complex environments.
2. At the model architecture level, the synergistic optimisation of C2f-ODConv, C2f-ACmix and ASFF-Head enhances the CCAI-YOLO model's feature extraction, global context modelling and multi-scale feature fusion capabilities, thereby strengthening detection performance of the model.
3. At the training optimisation level, the Inner-SIoU loss function is employed, integrating directional awareness with internal scaling. This enables prediction boxes to align rapidly and stably with target boxes, thereby enhancing model training efficiency.
4. Experimental results on the SSDD and SAR-Ship-Dataset datasets indicate that CCAI-YOLO exhibits enhanced accuracy and robustness compared to multiple alternative methods, thereby contributing to the advancement of maritime defence infrastructure.

The subsequent structure of this paper is as follows: Section 2 reviews relevant prior research; Section 3 elaborates on the proposed CCAI-YOLO model architecture and its design principles; Section 4 details the experimental setup and specific implementation methods; Section 5 analyses and discusses the experimental results; Section 6 systematically summarises the entire paper and draws conclusions.

2. Related Works

The YOLO series represents a cornerstone in object detection, having pioneered the single-stage paradigm that successfully balances speed with accuracy. As a mature and benchmark-setting iteration within this series, YOLOv8 remains a preferred choice in research due to its robust all-around performance in SAR image ship detection tasks.

To enhance the extraction and utilisation of multi-scale vessel features in SAR imagery, a series of improved models based on the YOLOv8 framework have been successively proposed. Regarding feature enhancement and multi-scale fusion, the MSFA-YOLO proposed by Zhao et al. [27] enhances multi-scale representations by integrating C2fSE and DenseASPP modules, though its robustness and accuracy remain subject to improvement. The DGSP-YOLO constructed by Zhu et al. [28] significantly enhances small object detection capability and noise resistance by embedding SPDConv, C2fMHSA, and DySample samplers. Regarding attention mechanisms and context modelling, Li et al. [29] proposed MAEE-Net, which integrates a Multi-Attention Feature Fusion Module (MAFM) and Edge Feature Enhancement Module (EFEM) at the neck to reinforce shallow target features while suppressing background interference. Luo et al. [30] designed SHIP-YOLO, incorporating a stochastic attention mechanism and Wise-IoU loss to address challenges posed by small targets and complex backgrounds. In terms of lightweight design and efficiency optimisation, Wang et al. [31] proposed YOLOSAR-Lite, which employs knowledge distillation and lightweight component replacement to reduce model complexity while maintaining accuracy. Regarding specialised detector heads and loss function design, the YOLOV8-FDF framework proposed by Jiang et al. [32] integrates the FADC module with a deformable feature adaptation mechanism, employing a dedicated detector head to enhance recognition accuracy for minute objects.

In summary, while existing studies have made notable advances in SAR ship detection by incorporating novel network modules, attention mechanisms, and refined loss functions, most efforts remain confined to localized architectural enhancements. These approaches typically focus on single performance metrics, lacking systematic design that addresses the synergistic relationship between feature characterisation capability, multi-scale contextual integration capability, and positioning accuracy. Particularly under typical SAR imaging conditions characterised by high noise, multi-scale targets, and complex sea–land backgrounds, existing architectures often fail to achieve lightweight designs while maintaining high detection precision. Consequently, models face severe challenges in real-world deployment scenarios, where balancing efficiency and effectiveness proves difficult.

To systematically address the aforementioned challenges, this study proposes CCAI-YOLO—a lightweight SAR ship detection framework based on a collaborative optimisation strategy. Unlike localised refinements, this approach comprehensively reconstructs YOLOv8's backbone network, neck structure, detection head, and loss function, with its design strictly addressing core difficulties in SAR image detection.

3. Materials and Methods

The overall architecture of the proposed CCAI-YOLO model is depicted in Figure 1. Building upon the YOLOv8 framework, this model employs a systematic design centred on architecture co-optimisation and supplemented by fine-tuned training strategies. It

aims to collaboratively address the specific challenges of ship detection in SAR imagery. At the model architecture level, we implemented three targeted enhancements constituting the core contributions of the algorithm: (1) Within the backbone network, the original C2f module on the key path was replaced with C2f-ODConv, leveraging dynamic convolutions to enhance adaptability and representational capacity for features exhibiting variable scales and orientations of ship targets; (2) Within the neck network, we introduced the C2f-ACmix module. Its parallel design of convolutional and self-attention operations strengthens global context modelling to suppress complex background noise; (3) We adopted the Adaptive Spatial Feature Fusion Detection Head (ASFF-Head), enabling adaptive weighted fusion of multi-scale features to enhance positioning accuracy for ships of varying dimensions. These three enhancements form the structural pillars underpinning the model's performance improvement. At the training optimisation level, we replaced the bounding box regression loss function with Inner-SIoU. This enhancement serves as a crucial auxiliary strategy, guiding the aforementioned optimised architecture to learn precise box regression parameters more efficiently through the introduction of direction-aware loss, thereby further improving model performance. Subsequent sections shall elaborate on the design principles of each module, and through systematic comparison and ablation experiments, shall, respectively, validate the contributions of architectural co-optimisation and loss function fine-tuning. Finally, the comprehensive impact of this research on the field of SAR ship detection shall be explored.

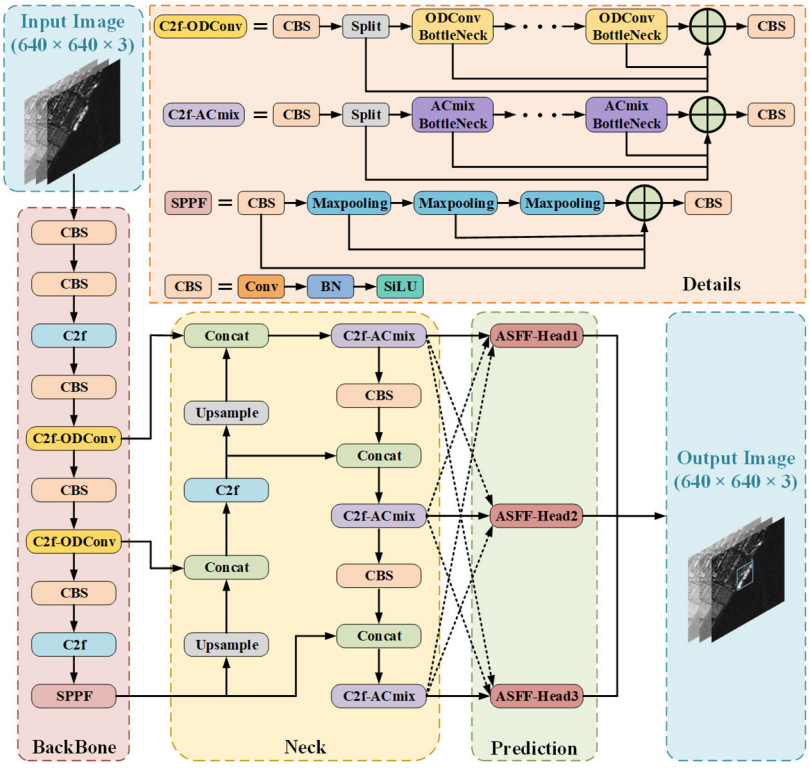


Figure 1. Overall structure of CCAI-YOLO.

3.1. C2f-ODConv Module

While prior dynamic convolution methods like CondConv [33] and DyConv [34] have been widely explored, they often share the same filter coefficients across each kernel, constraining their representational flexibility. These methods typically rely on linearly combining multiple static convolutions, which substantially increases parameter counts. In contrast, the Omni-Dimensional Dynamic Convolution (ODConv) [35] module adopted in this work evolves this concept by introducing a multi-dimensional attention mechanism

and a parallel strategy. ODConv performs linear weighting along four dimensions: the number of kernels, spatial size, input channel count and output channel count, enabling improved adaptation to complex SAR backgrounds and diverse ship shapes. Compared to earlier techniques, ODConv uses only a single convolutional kernel, significantly lowering parameters while maintaining high accuracy. By capturing inter-dimensional correlations, it enhances feature extraction capability and ensures efficient convolutional computation. The structure of ODConv is shown in Figure 2.

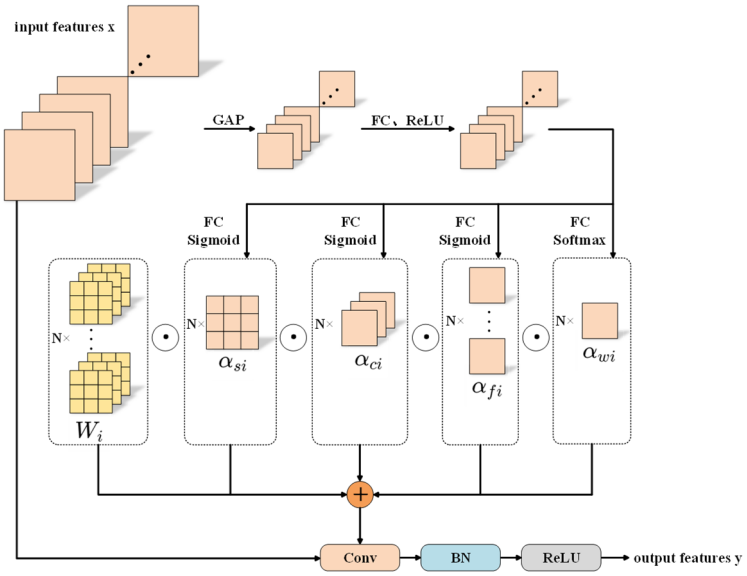


Figure 2. Details of ODConv module.

ODConv is expressed by Equation (1).

$$y = (\alpha_{\omega 1} \odot \alpha_{f1} \odot \alpha_{c1} \odot \alpha_{s1} \odot W_1 + \ldots + \alpha_{\omega n} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot W_n) * x \quad (1)$$

where $\alpha_{\omega i}$ represents the attention parameter of the convolution kernel W_i, α_{fi} represents the attention parameter of the output channel dimension, α_{ci} represents the attention parameter of the input channel dimension, and α_{si} represents the attention parameter of the convolution kernel spatial dimension, $\odot$ represents a weighted operation between convolutional filters of different dimensions.

Despite its advantages in lightweight design and gradient flow, the C2f module in YOLOv8 remains limited by its local receptive field when applied to SAR ship detection. Complex maritime backgrounds and sea clutter interference often lead to insufficient feature representation. Consequently, the C2f module struggles to fully capture the highly variable characteristics of the ship target. To address this issue, we propose the C2f-ODConv module, which replaces a CBS component in the C2f bottleneck with an ODConv layer. This targeted design aims to directly address the core feature constraints in SAR ship detection. The bottleneck layer of C2f adopts a 'compression-processing-expansion' architecture, serving as the core node for information filtering and enhancement within the feature flow. At the critical feature refinement stage—the second CBS component within the C2f bottleneck—ODConv is introduced to incorporate dynamicity. This replaces traditional static convolutions with input-dependent dynamic convolutional kernels, enabling the extraction of highly variable ship target features in SAR imagery. These features arise from noise, scale variations, and complex backgrounds. As illustrated in Figure 3, our enhancement mechanism strategically integrates ODConv's four-dimensional dynamic properties at this critical juncture. The spatial dynamic weighting mechanism adapts to local deformations and azimuth variations in ship targets caused by imaging geometry.

Channel dynamic attention autonomously enhances feature channels associated with strong scatter points on ships while suppressing channels contaminated by sea clutter or coherent speckle noise. The combined effects of filter dynamic combination and kernel size dynamic selection mechanisms collectively enhance the ability of the model to characterise the diverse scales and complex structures of ship targets. Therefore, C2f-ODConv is not merely an operator upgrade, but a purposeful design that embeds a dynamic perception mechanism at critical feature bottlenecks. It fundamentally enhances the feature extraction stage's ability to address specific challenges in SAR ship detection, improving detection accuracy while maintaining a lightweight module.

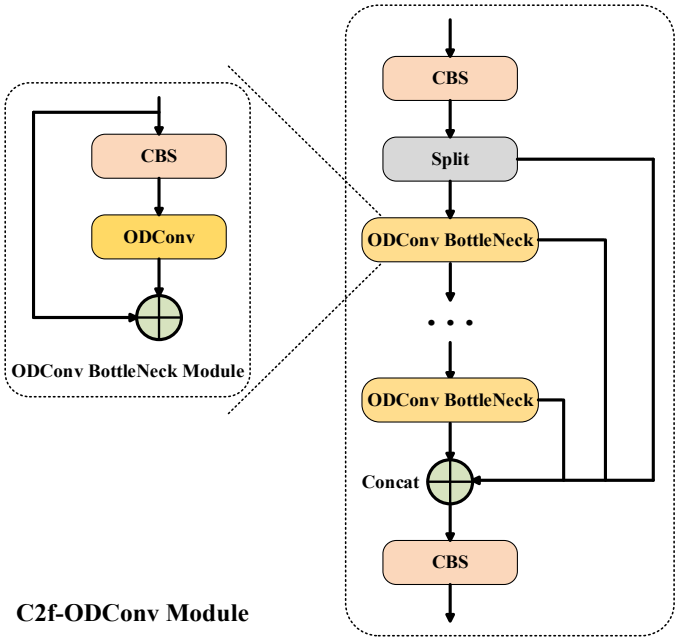


Figure 3. Details of C2f-ODConv module.

3.2. C2f-ACmix Module

Self-attention mechanisms play a crucial role in computer vision, particularly within the Transformer architecture [36], where they are extensively employed to model global dependencies within images. However, they often incur high computational costs. In contrast, convolutional operations efficiently extract local features such as edges and textures but lack global receptive fields. To bridge this gap, we adopt the ACmix module [37], a hybrid design that effectively integrates both paradigms. As shown in Figure 4, ACmix begins by projecting the input features into Query (Q), Key (K) and Value (V) tensors via three 1×1 convolutions. Q, K, V tensors are then processed along two parallel paths: a self-attention path, which reorganizes the tensors and computes attention-based aggregation, and a convolutional path, where the same tensors are reshaped and processed through fixed convolutional kernels. The outputs of both paths are summed, scaled by a learnable parameter, and combined with the input via a residual connection.

As shown in Figure 5, the ACmix module is seamlessly integrated into the C2f structure by placing it at the end of the Bottleneck, forming the C2f-ACmix module. This position is pivotal for achieving synergistic effects between local feature refinement and global semantic injection. Here, the local features extracted through prior convolutional layers possess foundational discriminative power, which the ACmix parallel dual-path mechanism strategically enhances. In addressing the specific challenges of SAR scenarios, this module enhances detection of local strong scatterers and edge structures on ships through its convolutional path. Concurrently, its self-attention path is dedicated to modelling long-

range semantic dependencies, effectively distinguishing systematic echoes from man-made regular structures such as harbours from the discrete strong scatter distribution of ship targets. By computing global correlations, it suppresses false alarms triggered by structured backgrounds. Positioning ACmix at the end of the C2f bottleneck layer, rather than at the beginning of the architecture or within the backbone network, represents a deliberate design trade-off. This ensures features undergo effective local abstraction and dimensionality reduction via front-end convolutions before incurring higher computational costs for global relationship modelling. By endowing features with critical contextual understanding just before output to the next stage at relatively economical computational expense, the module enhances robustness in complex scenarios while maintaining overall network efficiency.

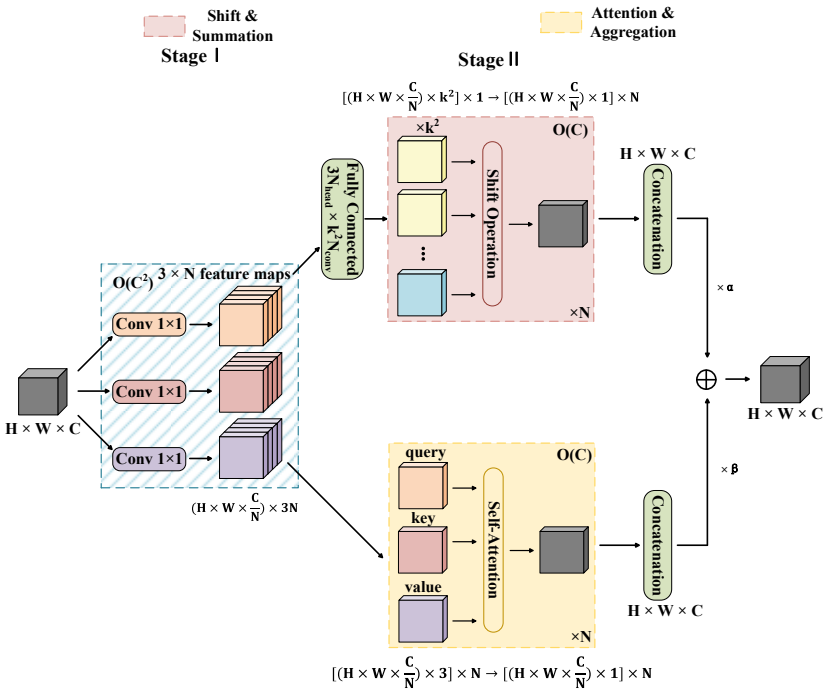


Figure 4. Details of ACmix attention.

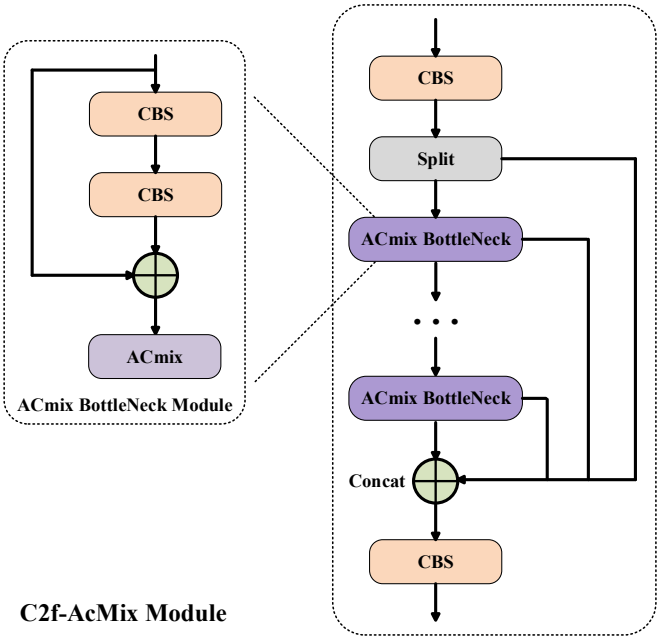


Figure 5. Details of C2f-ACmix module.

3.3. ASFF Detection Head

Ships in SAR imagery exhibit significant scale variations and often appear in dense clusters, leading to overlapping objects that complicate detection. The FPN-PAN structure in YOLOv8 also shows limited effectiveness in multi-scale feature fusion. To mitigate these limitations, we introduce the Adaptive Spatial Feature Fusion (ASFF) [38] detection head to replace the original YOLOv8 detection head, as shown in Figure 6.

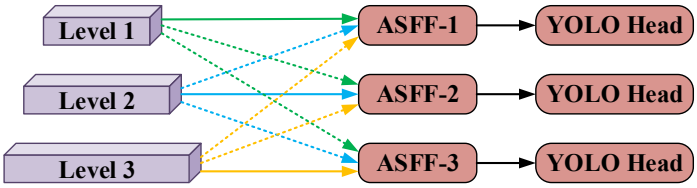


Figure 6. Details of ASFF detection Head.

The neck network of CCAI-YOLO outputs three multi-scale feature maps, denoted as level 1 to 3. Taking ASFF-1 as an example, its output is formed by adaptively fusing these features using learnable weights α, β, and γ, as defined in Equation (2):

$$ASFF-1 = \alpha^1 \cdot x^{1\to1} + \beta^1 \cdot x^{2\to1} + \gamma^1 \cdot x^{3\to1} \tag{2}$$

Here, x^1, x^2, and x^3 denote the input feature maps for level 1, level 2, and level 3, respectively. $x^{2\to1}$ represents upsampling the level 2 feature map to the same dimensions as level1, with $x^{3\to1}$ following the same principle. During the fusion process, the adjusted feature maps are multiplied by their corresponding adaptive weight parameters α^1, β^1 and γ^1 respectively. The weighted results are then summed to generate the final output feature map of the ASFF-1 module.

Let x denote the feature vector at position (i, j) on the feature map adjusted from layer n to layer l. The process of feature fusion at layer l is shown in Equation (3):

$$y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1\to l} + \beta_{ij}^l \cdot x_{ij}^{2\to l} + \gamma_{ij}^l \cdot x_{ij}^{3\to l} \tag{3}$$

Here, y_{ij}^l denotes the feature vector at position (i, j) within the output feature map y^l. α_{ij}^l, β_{ij}^l and γ_{ij}^l represent the spatial importance weights corresponding to the transmission of feature maps from three distinct hierarchical levels to the l layer; these weights are obtained through adaptive learning by the network. It should be noted that α_{ij}^l, β_{ij}^l and γ_{ij}^l are scalars and are shared across all channels. They satisfy the constraint: $\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l = 1$, $\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l \in [0, 1]$.

The specific calculation method is as follows:

$$\alpha_{ij}^l = \frac{e^{\lambda_{\alpha_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \tag{4}$$

Similarly, β_{ij}^l and γ_{ij}^l can be derived. Here, $\lambda_{\alpha_{ij}}^l$, $\lambda_{\beta_{ij}}^l$ and $\lambda_{\gamma_{ij}}^l$ serve as control parameters for the Softmax function. We compute these weight scalar maps using 1×1 convolutional layers, whose parameters can be learned through standard backpropagation. In this manner, features from each layer can be adaptively aggregated across different scales. The resulting fused features follow YOLO's transmission path to the detection head, ultimately serving for object classification and localisation.

In CCAI-YOLO, the original YOLOv8 detection head is replaced with an ASFF head. The core improvement of this module lies in its adaptive fusion of feature maps from

different scales within the neck network, achieved through learnable weight parameters prior to final predictions. Specifically, the ASFF mechanism dynamically assesses the reliability and importance of features at different levels across each spatial location. For SAR ship detection tasks, this implies the model autonomously determines greater reliance on deep, high-semantic features when detecting large ships near shore, while flexibly combining shallow, high-resolution detail features when capturing distant small targets. It generates clearer, more consistent multi-scale feature representations by suppressing responses from cluttered backgrounds or strong noise interference, whilst simultaneously enhancing semantic information highly relevant to ship targets.

3.4. Inner-SIoU Loss

Based on the Intersection over Union (IoU) metric, several advanced loss functions have been developed in recent years, including Generalized IoU (GIoU) [39], Distance IoU (DIoU) [40], Complete IoU (CIoU) [41], Enhanced IoU (EIoU) [42], SCYLLA-IoU (SIoU) [43] and Wise IoU (WIoU) [44]. GIoU addresses the zero-gradient issue in non-overlapping cases by incorporating the minimum enclosing box of both predicted and ground-truth bounding boxes. DIoU accelerates center alignment by penalizing the Euclidean distance between box centers. CIoU offers a more comprehensive optimization by considering overlap area, center distance, and aspect ratio. EIoU refines CIoU by reformulating the aspect ratio loss for more effective gradient propagation. SIoU introduces angle-based and shape-aware penalties to speed up convergence and improve accuracy. WIoU focuses on generalization by adapting to data quality. While YOLOv8 utilizes CIoU, it faces limitations in SAR ship detection due to the prevalence of low-quality training images. CIoU's aspect ratio term can over-penalize such samples, impairing generalization. Moreover, when center misalignment is large, the distance term may over-emphasize center matching at the expense of overlap optimization.

To overcome the convergence and scale sensitivity issues in bounding box regression for SAR imagery, we adopt the Inner-SIoU loss [45]. This loss function integrates the directional awareness mechanism of SIoU with the internal scaling concept of Inner-IoU. SIoU incorporates the vector angle between bounding boxes and redefines the penalty metric, while Inner-IoU introduces an internal scaling mechanism that calculates auxiliary IoU and corresponding penalty terms by simultaneously scaling both predicted and ground-truth boxes. The use of Inner-SIoU mitigates the performance degradation observed with CIoU loss function and speeds up convergence process, thereby boosting the adaptability of the model and generalization capabilities in dynamic detection environments. Figure 7 illustrates the Inner-IoU diagram, x_c^{gt} and y_c^{gt} denote the centre point coordinates of the ground truth box, x_c and y_c denote the centre point coordinates of the predicted box, w^{gt} and h^{gt} denote the width and height of the ground truth box, w and h denote the width and height of the predicted box, and *ratio* denotes the scaling factor for generating the auxiliary bounding box, typically ranging between [0.5, 1.5]. Among these, the scaling factor constitutes a core hyperparameter of this method. If the scaling factor is too large, the auxiliary box deviates insufficiently from the original box, resulting in inadequate constraints on centroid and size errors. Conversely, if the scaling factor is too small, the auxiliary box becomes excessively constricted, potentially causing training instability. Setting the scaling factor to 0.75 imposes a moderately tightening constraint on the predicted bounding box. Moreover, ship targets in SAR imagery typically exhibit relatively distinct scattering boundaries, though their appearance is affected by speckle noise. A scaling factor of 0.75 strengthens localisation constraints without unduly compressing the bounding box to the point of compromising the necessary tolerance for countering noise, thereby

achieving a favourable balance between enhanced accuracy and maintained robustness. Consequently, this value was consistently employed throughout all formal experiments.

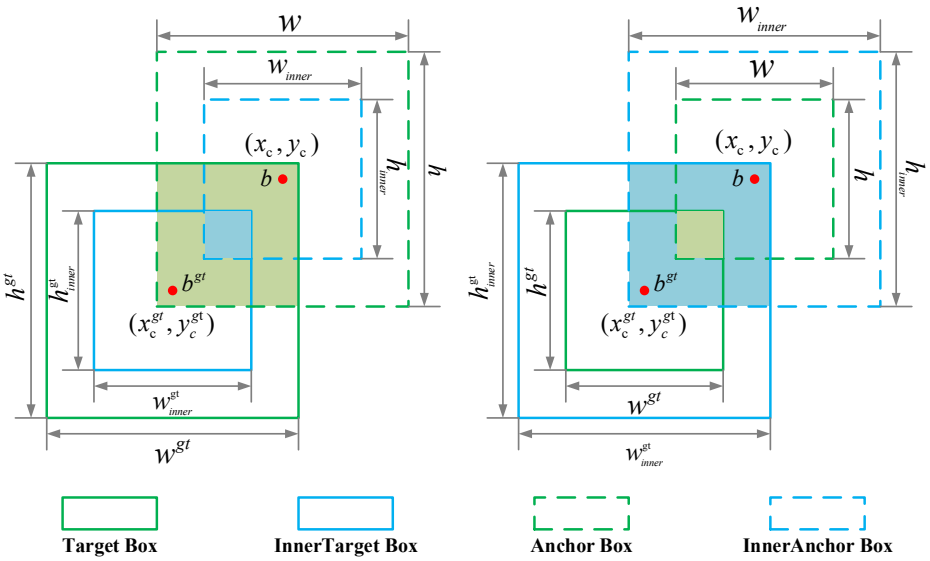


Figure 7. Schematic diagram of Inner-IoU.

Inner-SIoU combines Inner-IoU and SIoU, with its calculation formula as follows:

$$L_{Inner-SIoU} = L_{SIoU} + IoU - IoU^{inner} \tag{5}$$

where L_{SIOU} denotes the SIoU loss function, IoU denotes the intersection-over-union ratio between the predicted box and the ground truth box, and the definition of IoU^{inner} is shown in Equation (6).

$$IoU^{inner} = \frac{inter}{union} \tag{6}$$

The definitions of *inter* and *union* are shown in Equations (7) and (8):

$$inter = \left(min\left(b_r^{gt}, b_r\right) - max\left(b_l^{gt}, b_l\right)\right) * \left(min\left(b_b^{gt}, b_b\right) - max\left(b_t^{gt}, b_t\right)\right) \tag{7}$$

$$union = \left(w^{gt} * h^{gt}\right) * (ratio)^2 + (w * h) * (ratio)^2 - inter \tag{8}$$

where b_r^{gt}, b_r, b_l^{gt}, b_l, b_b^{gt}, b_b, b_t^{gt} and b_t are defined as follows.

$$b_r^{gt} = x_c^{gt} + \frac{w^{gt} * ratio}{2}, b_l^{gt} = x_c^{gt} - \frac{w^{gt} * ratio}{2} \tag{9}$$

$$b_b^{gt} = y_c^{gt} + \frac{h^{gt} * ratio}{2}, b_t^{gt} = y_c^{gt} - \frac{h^{gt} * ratio}{2} \tag{10}$$

$$b_r = x_c + \frac{w * ratio}{2}, b_l = x_c - \frac{w * ratio}{2} \tag{11}$$

$$b_b = y_c + \frac{h * ratio}{2}, b_t = y_c - \frac{h * ratio}{2} \tag{12}$$

Inner-IoU refines the evaluation of bounding box overlap by concentrating on their central regions. It incorporates a scale factor to adjust auxiliary box sizes, allowing flexible adaptation to different detection tasks and target types, thereby significantly boosting the model's adaptability and generalization. Meanwhile, SIoU improves convergence and localization accuracy by explicitly modeling directional relationships between boxes, as illustrated in Figure 8.

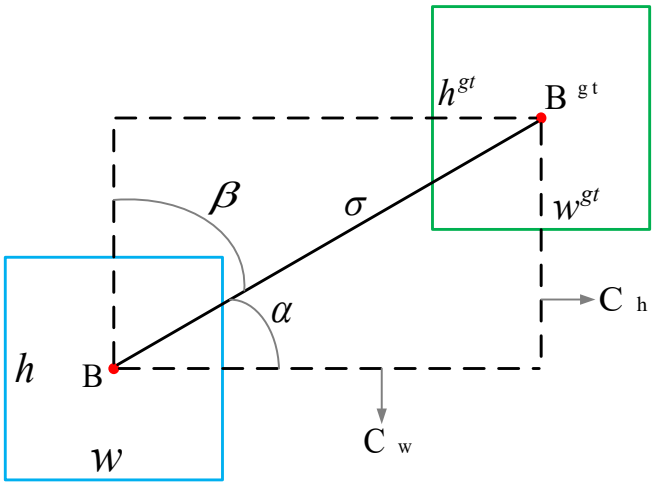


Figure 8. Schematic diagram of SIoU.

The SIoU loss calculation formula is as follows:

$$L_{SIoU} = 1 - IoU + \frac{(\Lambda + \Omega)}{2} \tag{13}$$

Here, Λ denotes the angular cost, where the angle in this cost function refers to the angle formed by the line connecting the centre point of the ground truth and the centre point of the predicted bounding box. The formula is as follows.

$$\Lambda = 1 - 2 \times \sin^2(\arcsin(\frac{c_h}{\sigma}) - \frac{\pi}{4}) \tag{14}$$

$$\sigma = \sqrt{\left(b_{cx}^{gt} - b_{cx}\right)^2 + \left(b_{cy}^{gt} - b_{cy}\right)^2} \tag{15}$$

$$c_h = \max(b_{cy}^{gt}, b_{cy}) - \min(b_{cy}^{gt}, b_{cy}) \tag{16}$$

where σ denotes the distance between the centres of the ground truth bounding box and the predicted bounding box, c_h represents the height difference between the centres of the ground truth and predicted bounding boxes, b_{cx}^{gt} and b_{cy}^{gt} denote the centre coordinates of the ground truth bounding box, and b_{cx} and b_{cy} denote the centre coordinates of the predicted bounding box.

Ω denotes shape cost, defined by the following formula.

$$\Omega = \sum_{t=w,h} (1 - e^{-wt})^{\theta} = (1 - e^{-w_w})^{\theta} + (1 - e^{-w_h})^{\theta} \tag{17}$$

$$w_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, \; w_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \tag{18}$$

w^{gt} and h^{gt} denote the width and height of the ground truth bounding box, w and h denote the width and height of the predicted bounding box, and θ represents the degree of concern for shape cost.

The Inner-SIoU loss function seamlessly integrates the strengths of both Inner-IoU and SIoU. This design offers a key advantage in SAR ship detection tasks: by improving the convergence dynamics of training, our proposed novel structural model can more efficiently learn precise localisation parameters. For ship targets occupying a small proportion of the image, minor positional deviations in initial prediction boxes can lead to substantial IoU loss. The angular constraints imposed by Inner-SIoU effectively reduce directional oscillations of prediction boxes during training, promoting faster and more stable alignment

with the target. Consequently, the Inner-SIoU loss function enhances the efficiency and stability of model training, thereby generating more reliable detection boxes in complex boundary scenarios.

4. Results

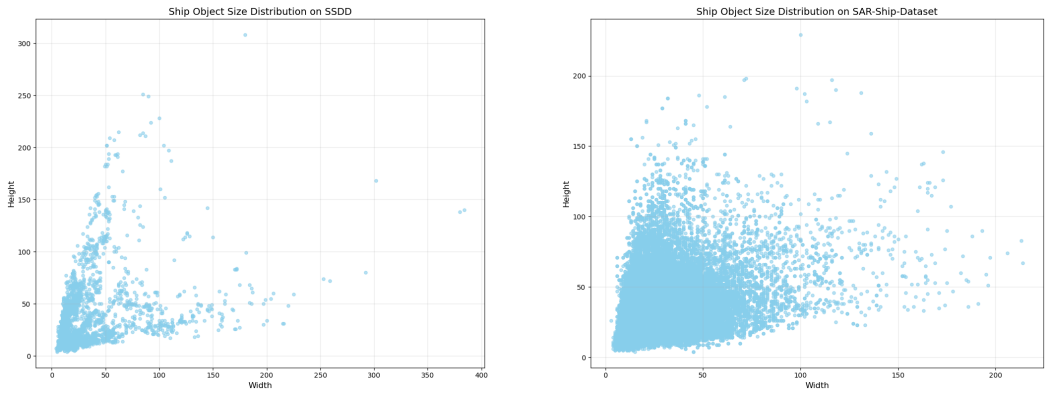
4.1. Datasets

To evaluate the proposed method, we employed two widely used public datasets: the SAR Ship Detection Dataset (SSDD) [46] and the SAR-Ship-Dataset [47]. The SSDD contains 1160 SAR images of various resolutions and 2,587 ship instances, covering diverse scenarios ranging from complex docks and coasts to open sea conditions. Following the official protocol, we split the data into training and validation sets at an 8:2 ratio, reserving the validation set images ending with digits '1' and '9' for testing. The model was trained for 300 epochs to ensure convergence. The SAR-Ship-Dataset consists of 39,729 SAR images at 256 × 256 pixels comprising 50,885 ship targets. It exhibits significant variations in scale and orientation, with each image carefully deduplicated to ensure diversity and accuracy. Adhering to the standard partition strategy, we divided the dataset randomly into training, validation, and test sets at a 7:2:1 ratio for fair comparison with existing methods. Due to the larger volume of data, the training was extended to 400 epochs. The parameter settings and data partitioning information for both datasets are summarized in Table 1.

Table 1. Parameter configuration for the dataset.

Datasets	SSDD	SAR-Ship-Dataset
Number of images	1160	39,729
Number of ships	2587	50,885
Number of classes	1	1
Image size	About 500 × 500	256 × 256
Dataset splitting	8:2	7:2:1
Epochs	300	400

The size distribution of ship targets plays a crucial role in the design and evaluation of detection algorithms. Based on the SSDD and SAR-Ship-Dataset, we visualize the width and height distributions of ship targets, as shown in Figure 9.



(**a**) (**b**)

Figure 9. Distribution of width and height across two datasets. (**a**) SSDD. (**b**) SAR-Ship-Dataset.

Figure 9a reveals that in SSDD, most ship targets have widths and heights concentrated in the 0–100 pixel range, indicating a predominance of small-sized ships. Meanwhile, some targets exceed 100 pixels, with a very small number surpassing 300 pixels, reflecting a

broad multi-scale size distribution in the dataset. This wide size variation poses challenges for detection algorithms. Small targets, with limited pixels and weak feature expression, are prone to being missed or misclassified. Although large targets provide richer features, they require a balance between local details and global structure, placing higher demands on the model's ability to represent and recognize objects across scales. As shown in Figure 9b, the SAR-Ship-Dataset exhibits a similar concentration of ship sizes within the 0–100 pixel range. This distribution highlights that small target detection is the central challenge in this dataset, requiring algorithms with strong capabilities in capturing and discriminating subtle features of small objects.

4.2. Experimental Settings

All experiments were conducted within the same software and hardware environment to ensure comparability and reproducibility of results. The specific configuration is as follows: the hardware platform comprised a workstation equipped with an Intel® Core™ i9-11900K@3.50 GHz processor, 16 GB of memory, and an NVIDIA GeForce RTX 3090 graphics card, running the Ubuntu 18.04.6 LTS operating system. The training code was developed using Python 3.10.15 and the PyTorch 2.0.0 framework, with GPU acceleration enabled via CUDA 11.7. Regarding key hyperparameters and training strategies for the model, we employed the following settings: input images were uniformly resized to 640 × 640 pixels, with a batch size of 16. The Adam optimiser was utilised, featuring an initial learning rate of 0.002 and a momentum parameter of 0.9. Weight decay (coefficient 0.0005) was applied to mitigate overfitting. Model weights were randomly initialised; no externally pre-trained weights were utilised. The number of worker threads for data loading was set to 8. The software and hardware environments required for the experiments, along with relevant parameter settings, are detailed in Tables 2 and 3, respectively.

Table 2. Experimental software and hardware environment.

Item	Configuration
CPU	Intel® Core™ i9-11900K
GPU	NVIDIA GeForce RTX 3090
RAM	16 GB
Operating System	Ubuntu 18.04.6 LTS
Programming Language	Python 3.10.15
Algorithm Framework	Pytorch 2.0.0

Table 3. Experimental parameter settings.

Name	Value
Image size	640 × 640
Batch size	16
Workers	8
Pretraining weight	0
Optimizer	Adam
Learning rate	0.002
Weight decay	0.0005
Momentum	0.9

4.3. Evaluation Metrics

Precision (P) and recall (R) are fundamental metrics in target detection. Precision reflects the accuracy of positive predictions, while recall measures the completeness of true

positive detections. Their calculations are defined in Equations (19) and (20), where TP, FP, and FN denote true positives, false positives, and false negatives, respectively.

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{19}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{20}$$

Since precision and recall often exhibit a trade-off, the F1 score—their harmonic mean—is widely adopted as a balanced performance measure, as shown in Equation (21). A higher F1 score indicates a better balance between the two metrics.

$$\mathrm{F1} = 2 \times \frac{\mathrm{P} \times \mathrm{R}}{\mathrm{P} \times \mathrm{R}} \tag{21}$$

Mean Average Precision (mAP) is the most authoritative comprehensive metric in the field of target detection, representing the average performance across all categories. As this study detects only ship-type targets, mAP is equivalent to Average Precision (AP), whose calculation formula is shown in Equation (22).

$$\mathrm{mAP} = \frac{1}{N} \sum_{i=1}^{N} \int_0^1 P(\mathrm{R}) d\mathrm{R} = \int_0^1 P(\mathrm{R}) d\mathrm{R} \tag{22}$$

4.4. Ablation Experiments

In this study, YOLOv8n is selected as the baseline model, and a series of ablation experiments are conducted on the SSDD and SAR-Ship-Dataset to validate the effectiveness of the proposed improvements. Six key metrics were adopted for comprehensive evaluation: Precision (P), Recall (R), mAP50, mAP50-95, parameter count and computational complexity. Among them, mAP50 is considered critical indicator, reflecting overall detection capability of the model under a relaxed IoU threshold. Concurrently, mAP50-95—as the average accuracy across IoU thresholds from 50% to 95%—enables a more comprehensive assessment of a model's robustness under varying precision requirements. This metric proves particularly valuable in applications demanding high bounding box alignment accuracy.

As shown in Table 4, ablation experiments systematically evaluated the impact of each module on accuracy, efficiency and multi-scale performance. Integrating C2f-ODConv into the baseline model moderately improved *p*-value, R-value, mAP50 and mAP50 while increasing parameters by only approximately 0.287 million and reducing FLOPs by 0.5 billion. This demonstrates that the dynamic convolution mechanism of C2f-ODConv not only captures multi-scale features more effectively but also achieves lightweight operation through optimised computational pathways. When C2f-AcMix is introduced independently, it achieves a 2.8% improvement in *p*-value and a 1.4% increase in mAP50-95 on the SSDD, though the R-value decreases by 1.6%. Computational cost (FLOPs) increases by merely 0.2 G. The decline in recall rate stems from the module's enhanced global context modelling through self-attention, which improves its ability to distinguish complex structured backgrounds such as port facilities. This heightened discriminative capability may temporarily filter out certain genuinely small targets with extremely faint or highly obscured features during the initial training phase, resulting in a slight increase in missed detections. The addition of the ASFF head module incurs the most significant structural overhead (approximately 1.373 million additional parameters and 2.2 GFLOPs), yet it substantially improves both *p*-value and mAP50 by optimising multi-scale fusion, demonstrating its efficacy in complex scenarios such as overlapping ships. Employing the Inner-SIoU loss function near-universal enhances all four metrics without altering any model parameters

or computational complexity, validating its value as an efficient optimisation strategy. When all modules operate in concert, CCAI-YOLO achieves optimal accuracy on both the SSDD and SAR-Ship-Dataset (e.g., mAP50-95 reaching 74.1% and 71.6%, respectively, with mAP50 at 98.8% and 98.2%), with a total parameter count of 4.754 million and FLOPs of 10.1 billion. Comprehensive analysis indicates that CCAI-YOLO achieves a more significant improvement in detection accuracy at the cost of approximately 57.8% more parameters and 23.2% higher computational complexity compared to baseline models. The CCAI-YOLO model achieves a competitive balance between accuracy and efficiency, rendering it suitable for SAR ship detection scenarios demanding stringent accuracy requirements alongside moderate computational resources.

Table 4. Ablation experiments on SSDD and SAR-Ship-Dataset.

Baseline	C2f-ODConv	C2f-AcMix	ASFF Head	Inner-SIoU	Params (M)	FLOPs (G)	SSDD				SAR-Ship-Dataset			
							P	R	mAP50	mAP50-95	P	R	mAP50	mAP50-95
√					3.011	8.2	0.949	0.955	0.980	0.722	0.954	0.947	0.978	0.702
√	√				3.298	7.7	0.955	0.961	0.985	0.729	0.958	0.952	0.980	0.706
√		√			3.095	8.4	0.977	0.939	0.987	0.736	0.951	0.948	0.977	0.704
√			√		4.384	10.4	0.973	0.941	0.982	0.725	0.957	0.952	0.979	0.709
√				√	3.011	8.2	0.956	0.962	0.985	0.723	0.955	0.948	0.979	0.703
√	√	√			3.382	7.9	0.972	0.958	0.987	0.734	0.957	0.951	0.978	0.710
√	√	√	√		4.754	10.1	0.974	0.961	0.986	0.736	0.958	0.955	0.980	0.713
√	√	√	√	√	4.754	10.1	**0.980**	**0.969**	**0.988**	**0.741**	**0.960**	**0.956**	**0.982**	**0.716**

Note: The symbol '√' indicates that the module has been added to the model. The best results are in bold.

4.5. Comparative Experiment

To further evaluate the effectiveness of the proposed CCAI-YOLO model, comparative experiments were carried out on the SSDD and SAR-Ship-Dataset against several lightweight YOLO-family models—YOLOv8n, YOLOv9t, YOLOv10n, and YOLOv11n—as well as recently published YOLO-based detectors such as MSFA-YOLO, MAEE-Net, YOLOSAR-Lite, SHIP-YOLO, YOLO-SRBD [48], SwinYOLOv7 [49] and CSD-YOLO [50]. To ensure fairness in comparisons, all baseline models from the YOLO series were retrained under identical experimental conditions (input size 640 × 640 pixels, employing uniform hyperparameters and data augmentation strategies). For other models under comparison, we directly referenced the performance metrics reported in their original papers. In addition to employing the six evaluation metrics outlined in Section 4.4, we have incorporated the F1 score to comprehensively assess the model's performance. Comparative results are summarized in Tables 5 and 6, with best results highlighted in bold. Entries marked as "-" indicate unreported values in the original references.

Based on the comprehensive comparison results presented in Table 5, CCAI-YOLO demonstrates outstanding overall performance in SSDD. It achieves the highest p (0.980), mAP50 (0.988) and mAP50-95 (0.749), while its F1 score (0.973) is merely 0.005 lower than that of the top-performing MSFA-YOLO model. Compared to the baseline model YOLOv8n, CCAI-YOLO achieves an 0.8 percentage point improvement in mAP50 at the cost of approximately 1.743 million additional parameters and 1.9 GFLOPs of computational overhead. Notably, CCAI-YOLO achieved a significant lead on the more challenging mAP50-95 metric, outperforming YOLOv8n by approximately 2.7 percentage points. demonstrating its ability to maintain highly stable detection performance under stricter localisation requirements. Further analysis reveals our model excels in balancing efficiency and accuracy: compared to the ultra-lightweight YOLOv9t model with merely 1.765 million parameters, CCAI-YOLO achieves a 1.1 percentage point lead in mAP50 and a 2.9 percentage point lead in mAP50-95. while outperforming certain specialised models with substantially increased parameters (such as SwinYOLOv7 with 32.75 million parameters) by achieving higher accuracy with fewer than 15% of the parameters. Notably, when benchmarked against other lightweight

variants, CCAI-YOLO achieves superior F1 scores and mAP50 compared to YOLOSAR-Lite and SHIP-YOLO while maintaining comparable parameter counts.

Table 5. Comparison of different algorithms on SSDD.

Models	P	R	F1	mAP50	mAP50-95	Params (M)	FLOPs (G)
YOLOv8n	0.949	0.955	0.952	0.980	0.722	3.011	8.2
YOLOv9t	0.965	0.932	0.948	0.977	0.720	1.765	6.7
YOLOv10n	0.925	0.925	0.925	0.967	0.715	2.707	8.4
YOLOv11n	0.949	0.954	0.951	0.978	0.711	2.59	6.4
MSFA-YOLO	0.977	**0.980**	**0.978**	0.987	0.662	-	-
MAEE-Net	0.975	0.934	0.954	0.986	-	16.2	-
YOLOSAR-Lite	0.923	0.912	0.918	0.953	-	2.05	4.48
SHIP-YOLO	0.971	0.945	0.958	0.971	-	2.5	7.2
YOLO-SRBD	0.979	0.960	0.969	0.983	0.745	7.72	21.5
SwinYOLOv7	0.946	0.893	0.911	0.966	-	32.75	112.57
CSD-YOLO	0.959	0.959	0.959	0.986	0.691	-	-
Ours	**0.980**	0.969	0.973	**0.988**	**0.749**	4.754	10.1

Note: To facilitate comparative analysis, the optimal values for each metric in the table are indicated in bold font.

Table 6. Comparison of different algorithms on SAR-Ship-Dataset.

Models	P	R	F1	mAP50	mAP50-95	Params (M)	FLOPs (G)
YOLOv8n	0.954	0.947	0.950	0.978	0.702	3.011	8.2
YOLOv9t	0.944	0.942	0.943	0.975	0.692	1.765	6.7
YOLOv10n	0.946	0.947	0.946	0.975	0.699	2.707	8.4
YOLOv11n	0.949	0.951	0.950	0.979	0.702	2.59	6.4
MSFA-YOLO	0.945	0.933	0.939	0.969	0.677	-	-
MAEE-Net	0.928	0.910	0.919	0.947	-	16.2	-
YOLOSAR-Lite	0.927	0.915	0.921	0.961	-	2.05	4.48
SHIP-YOLO	0.932	0.928	0.930	0.966	-	2.5	7.2
Ours	**0.960**	**0.956**	**0.958**	**0.982**	**0.716**	4.754	10.1

Note: To facilitate comparative analysis, the optimal values for each metric in the table are indicated in bold font.

Table 6 presents the results on the SAR-Ship-Dataset, where CCAI-YOLO again achieves the best F1 score (0.958), mAP50 (0.982) and mAP50-95 (0.716), improving over YOLOv8n by 0.008, 0.004 and 0.014, respectively. With 4.754M parameters, the model remains considerably lighter than larger detectors such as MAEE-Net (16.2 M), demonstrating a favorable trade-off between accuracy and complexity suitable for edge deployment.

In summary, comparative experiments demonstrate that CCAI-YOLO does not enhance performance through the simple stacking of parameters, but rather achieves effective in-novation in its model architecture. It strikes a favourable balance across multiple key dimensions, including accuracy (particularly mAP50, mAP50-95 and precision), efficiency (parameter count and computational load), and robustness (F1 score). Its design finds a competitive equilibrium between lightweight implementation and high precision.

4.6. Visualisation of Test Results

To demonstrate the effectiveness of our method, we selected representative samples from the SSDD and SAR-Ship-Dataset for visual comparison. Detection results are shown in Figure 10. From left to right, each group presents: ground truth, YOLOv8n, YOLOv11n, and our model. In the visualizations, green boxes indicate ground truth annotations, sky blue boxes mark correct detections, red ellipses highlight false positives and yellow ellipses indicate missed detections.

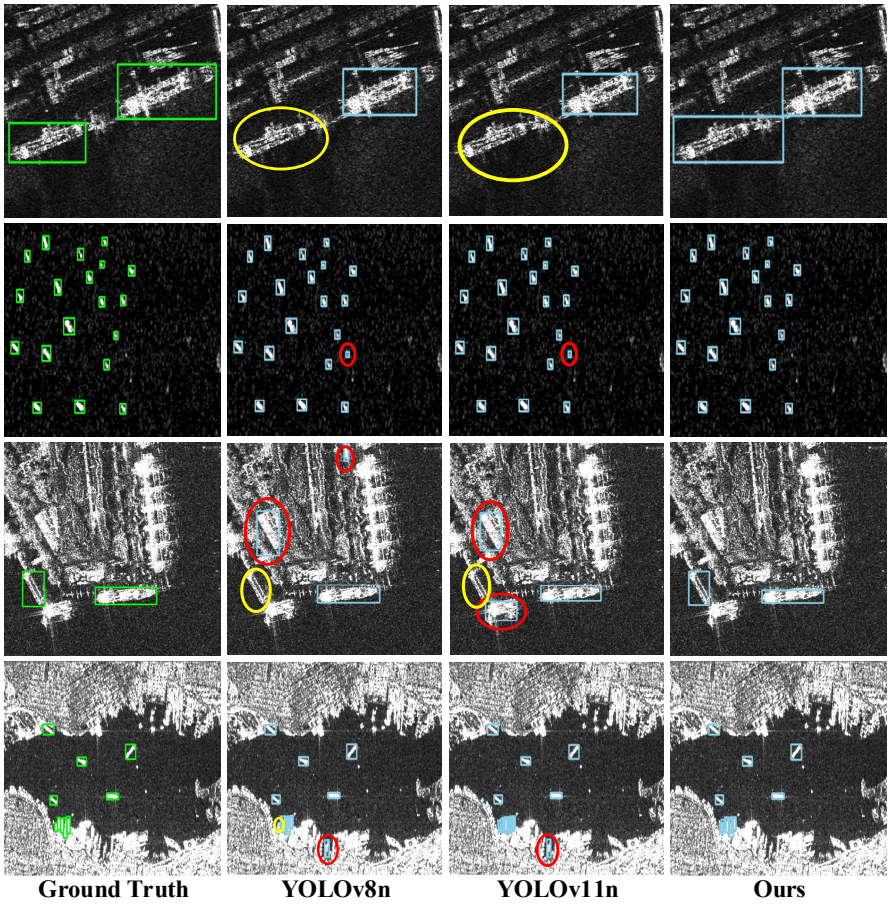


Figure 10. Ship detection results on SSDD.

As shown in Figure 10, the visualisation results on the SSDD concretely demonstrate the performance gains achieved through enhancements to each module of CCAI-YOLO. In the first sample, both YOLOv8n and YOLOv11n failed to detect medium-sized ships near the coastline. CCAI-YOLO, however, achieved robust recognition of all targets through its enhanced feature extraction capability enabled by the C2f-ODConv module. In the second sample, which depicts a distant coastal scene with small-sized ships, the baseline models produced false alarms due to the visual similarity between islands and ships. In contrast, CCAI-YOLO effectively distinguished semantic differences between background and targets through global context modelling via the C2f-ACmix module, achieving precise localisation without false alarms. The third and fourth samples, captured in noisy near-shore environments with heavy interference from islands and land structures, further reveal the limitations of YOLOv8n and YOLOv11n, where multiple false positives and missed detections occurred. CCAI-YOLO enhanced the consistency of target representation through the adaptive fusion of multi-scale features via ASFF-Head, while the attention mechanism of C2f-ACmix effectively suppresses background interference. Consequently, it maintains high accuracy and zero false alarms even in highly noisy environments.

Figure 11 presents visualisation results on the more challenging SAR-Ship-Dataset, which is characterized by strong speckle noise, large scale variations, and diverse maritime scenes. The first three samples, drawn from complex coastal environments, triggered numerous false detections in YOLOv8n and YOLOv11n. Our model, by comparison, maintained high detection accuracy and stability. In the fourth sample, which includes a distant shoreline, our approach reliably distinguished ships from the coastal background, demonstrating superior generalization and robustness.

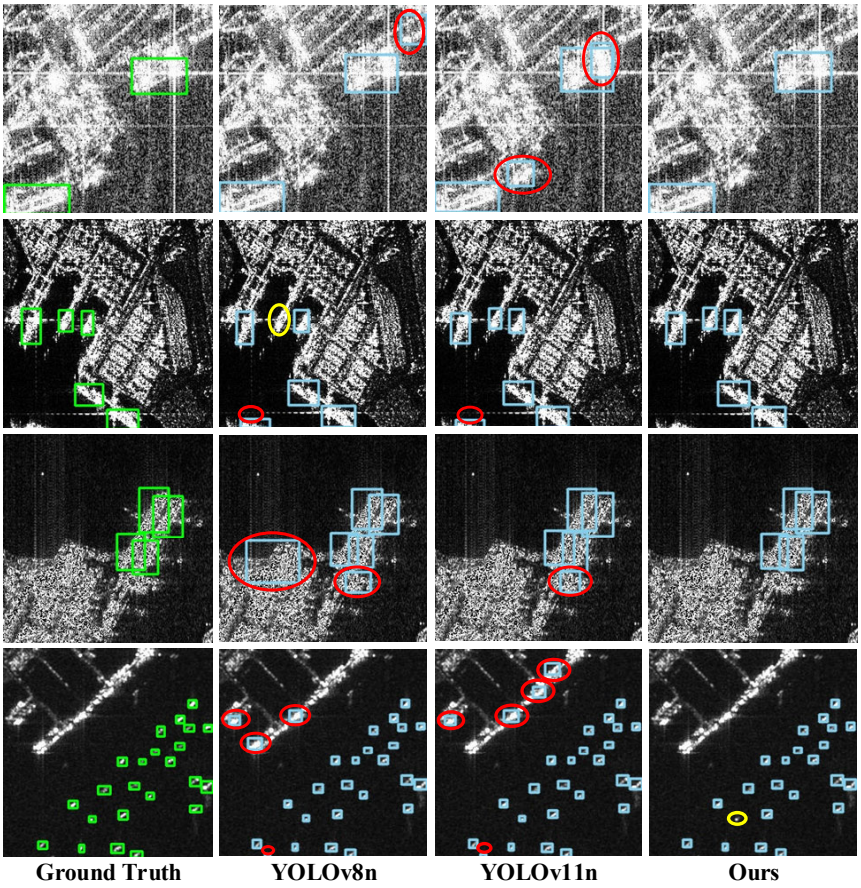


Figure 11. Ship detection results on SAR-Ship-Dataset.

Target detection models, despite their formidable performance, often lack interpretability in their internal decision-making mechanisms, making it difficult to intuitively understand the basis for their object detection decisions. Gradient-Weighted Class Activation Maps (Grad-CAM) [51] effectively address this issue by generating heatmaps. Within these heatmaps, deep red indicates the highest level of model attention to a region, while blue areas denote lower levels of attention.

To compare models' interpretability, we selected four typical scenarios from two datasets for heatmap visualisation analysis, with results shown in Figure 12. Each image group displays, from left to right: ground truth, YOLOv8n heatmap results, and our proposed CCAI-YOLO model's heatmap results. In the first two comparisons using the SSDD, it is evident that in offshore scenarios, CCAI-YOLO demonstrates heightened attention towards medium-sized ships, with its heatmap's red regions nearly fully covering the hulls. Conversely, in coastal scenarios with dense ship clusters, YOLOv8n exhibits missed detections for partially overlapping hulls, whereas CCAI-YOLO maintains robust detection capability, significantly reducing the false negative rate. In the latter two comparisons on the SAR-Ship-Dataset, CCAI-YOLO shows superior recognition of small targets and effectively suppresses false activations near islands or complex coastlines, confirming its stronger generalization capacity.

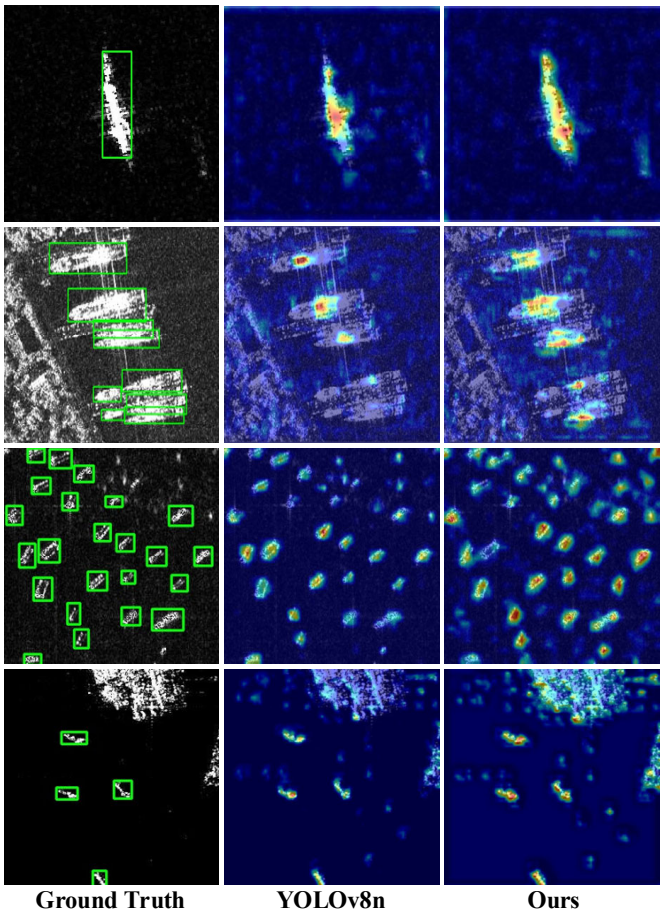


Figure 12. Grad-CAM visualization results of SSDD and SAR-Ship-Dataset.

5. Discussion

Based on the results from ablation experiments, comparative studies and detection visualizations, CCAI-YOLO exhibits strong adaptability for ship detection across diverse scenarios. This capability originates from the coordinated design of the four proposed enhancements, which collectively address characteristic challenges in SAR imagery—including inherent speckle noise, complex near-shore backgrounds with land and infrastructure and multi-scale ship distribution.

Although the model achieves promising results under experimental conditions, its performance remains to be verified in real-world spaceborne SAR environments. Further assessment is needed to determine its stability and robustness in operational settings. Future work will emphasize evaluating the model under authentic space mission conditions and developing advanced data augmentation techniques to improve both SAR image interpretability and generalization. Moreover, the model requires further lightweight optimization to minimize computational demands, facilitating deployment across a broader range of satellite platforms and promoting practical adoption in maritime remote sensing applications.

6. Conclusions

This paper presents CCAI-YOLO, an improved high-precision model for ship detection in SAR images, based on YOLOv8n. By integrating the C2f-ODConv module into the backbone network, the model effectively handles high noise and complex backgrounds in SAR imagery, enabling more adaptive feature extraction. The inclusion of the C2f-ACmix module in the neck enhances the capture of global contextual information, improving

ship localization and recognition accuracy while maintaining computational efficiency. The detection head employs an ASFF architecture to mitigate inconsistencies arising from multi-scale feature fusion. Furthermore, Inner-SIoU loss function has been incorporated, designed to enhance the model's convergence and localisation capability. On the SSDD and SAR-Ship-Dataset, CCAI-YOLO achieved F1 scores of 0.973 and 0.958, respectively, mAP50 values of 0.988 and 0.982, and mAP50-95 values of 0.749 and 0.716, demonstrating leading overall performance.

However, with ongoing advances in remote sensing sensors, processing high-resolution SAR imagery has become an inevitable trend. In real-world deployment, the real-time performance and efficiency of detection algorithms remain critical challenges. Achieving a lightweight model without sacrificing accuracy is still a core issue to be addressed. Furthermore, conventional axis-aligned bounding boxes struggle to handle overlapping objects effectively. The introduction of rotated bounding boxes could provide azimuth information of ship targets, enabling more precise separation from the background and improving localization accuracy. Techniques such as knowledge distillation and network pruning also offer promising avenues for model optimization. In future work, we will focus on further refining the model architecture to improve its generalization and robustness in complex real-world scenarios, keeping pace with evolving application requirements.

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

Funding: This research was funded by the National Natural Science Foundation of China, grant number 61501019; the Cultivation Project Funds for Beijing University of Civil Engineering and Architecture, grant number X24031.

Data Availability Statement: This study utilized two publicly available datasets: the SSDD and the SAR-Ship-Dataset. The data sources are accessible at: https://github.com/TianwenZhang0825/Official-SSDD and https://github.com/CAESAR-Radi/SAR-Ship-Dataset (both accessed on 1 January 2025).

Acknowledgments: The authors wish to express their sincere appreciation to the teams behind the SSDD and SAR-Ship-Dataset. Their significant contributions in building these meticulously annotated datasets were invaluable to this work.

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

References

1. Wang, C.; Jiang, S.; Zhang, H.; Wu, F.; Zhang, B. Ship Detection for High-Resolution SAR Images Based on Feature Analysis. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 119–123. [CrossRef]
2. Tang, J.; Deng, C.; Huang, G.-B.; Zhao, B. Compressed-Domain Ship Detection on Spaceborne Optical Image Using Deep Neural Network and Extreme Learning Machine. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1174–1185. [CrossRef]
3. Qi, S.; Ma, J.; Lin, J.; Li, Y.; Tian, J. Unsupervised Ship Detection Based on Saliency and S-HOG Descriptor From Optical Satellite Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1451–1455. [CrossRef]
4. Zhang, C.; Liu, P.; Wang, H.; Jin, Y. A review of recent advance of ship detection in single-channel SAR images. *Waves Random Complex Media* **2023**, *33*, 1442–1473. [CrossRef]
5. Schou, J.; Skriver, H.; Nielsen, A.A.; Conradsen, K. CFAR edge detector for polarimetric SAR images. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 20–32. [CrossRef]
6. Zhu, C.; Zhou, H.; Wang, R.; Guo, J. A Novel Hierarchical Method of Ship Detection from Spaceborne Optical Image Based on Shape and Texture Features. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3446–3456. [CrossRef]

7. Leng, X.; Ji, K.; Zhou, S.; Xing, X.; Zou, H. An Adaptive Ship Detection Scheme for Spaceborne SAR Imagery. *Sensors* **2016**, *16*, 1345. [CrossRef] [PubMed]
8. Tello, M.; Lopez-Martinez, C.; Mallorqui, J.J. A Novel Algorithm for Ship Detection in SAR Imagery Based on the Wavelet Transform. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 201–205. [CrossRef]
9. Ai, J.; Yang, X.; Song, J.; Dong, Z.; Jia, L.; Zhou, F. An Adaptively Truncated Clutter-Statistics-Based Two-Parameter CFAR Detector in SAR Imagery. *IEEE J. Ocean. Eng.* **2018**, *43*, 267–279. [CrossRef]
10. Ao, W.; Xu, F.; Li, Y.; Wang, H. Detection and Discrimination of Ship Targets in Complex Background From Spaceborne ALOS-2 SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 536–550. [CrossRef]
11. Chen, S.; Li, X. A new CFAR algorithm based on variable window for ship target detection in SAR images. *Signal Image Video Process.* **2019**, *13*, 779–786. [CrossRef]
12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*; NeurIPS Proceedings: San Diego, CA, USA, 2012.
13. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. *arXiv* **2013**, arXiv:1311.2524.
14. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef] [PubMed]
15. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into High Quality Object Detection. *arXiv* **2017**, arXiv:1712.00726. [CrossRef]
16. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 386–397. [CrossRef]
17. Xiao, Q.; Cheng, Y.; Xiao, M.; Zhang, J.; Shi, H.; Niu, L.; Ge, C.; Lang, H. Improved region convolutional neural network for ship detection in multiresolution synthetic aperture radar images. *Concurr. Comput. Pract. Exp.* **2020**, *25*, e5820. [CrossRef]
18. Ke, X.; Zhang, X.; Zhang, T.; Shi, J.; Wei, S. SAR ship detection based on an improved faster R-CNN using deformable convolution. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 3565–3568.
19. Jian, L.; Pu, Z.; Zhu, L.; Yao, T.; Liang, X. SS R-CNN: Self-Supervised Learning Improving Mask R-CNN for Ship Detection in Remote Sensing Images. *Remote Sens.* **2022**, *14*, 4383. [CrossRef]
20. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2015**, arXiv:1506.02640.
21. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. *arXiv* **2015**, arXiv:1512.02325.
22. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 318–327. [CrossRef]
23. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. *arXiv* **2019**, arXiv:1904.01355. [CrossRef]
24. Yu, C.; Shin, Y. SAR ship detection based on improved YOLOv5 and BiFPN. *ICT Express* **2023**, *10*, 28–33. [CrossRef]
25. Miao, T.; Zeng, H.; Wang, H.; Yang, W. Inshore ship detection in SAR images via an improved SSD model with wavelet decomposition. In Proceedings of the 2021 7th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), Bali, Indonesia, 1–3 November 2021; pp. 1–5.
26. Yang, S.; An, W.; Li, S.; Wei, G.; Zou, B. An Improved FCOS Method for Ship Detection in SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8910–8927. [CrossRef]
27. Zhao, L.; Ning, F.; Xi, Y.; Liang, G.; He, Z.; Zhang, Y. MSFA-YOLO: A Multi-Scale SAR Ship Detection Algorithm Based on Fused Attention. *IEEE Access* **2024**, *12*, 24554–24568. [CrossRef]
28. Zhu, L.; Chen, J.; Chen, J.; Yang, H. DGSP-YOLO: A Novel High-Precision Synthetic Aperture Radar (SAR) Ship Detection Model. *IEEE Access* **2024**, *12*, 167919–167933. [CrossRef]
29. Li, Z.; Ma, H.; Guo, Z. MAEE-Net: SAR ship target detection network based on multi-input attention and edge feature enhancement. *Digit. Signal Process.* **2024**, *156*, 104810. [CrossRef]
30. Luo, Y.; Li, M.; Wen, G.; Tan, Y.; Shi, C. SHIP-YOLO: A Lightweight Synthetic Aperture Radar Ship Detection Model Based on YOLOv8n Algorithm. *IEEE Access* **2024**, *12*, 37030–37041. [CrossRef]
31. Wang, H.; Shi, J.; Karimian, H.; Liu, F.; Wang, F. YOLOSAR-Lite: A lightweight framework for real-time ship detection in SAR imagery. *Int. J. Digit. Earth* **2024**, *17*, 2405525. [CrossRef]
32. Jiang, W.; Han, D.; Han, B.; Wu, Z. YOLOv8-FDF: A Small Target Detection Algorithm in Complex Scenes. *IEEE Access* **2024**, *12*, 119223–119237. [CrossRef]
33. Yang, B.; Bender, G.; Le, Q.V.; Ngiam, J. CondConv: Conditionally Parameterized Convolutions for Efficient Inference. *arXiv* **2019**, arXiv:1904.04971.
34. Chen, Y.; Dai, X.; Liu, M.; Chen, D.; Yuan, L.; Liu, Z. Dynamic Convolution: Attention over Convolution Kernels. *arXiv* **2019**, arXiv:1912.03458.

35. Li, C.; Zhou, A.; Yao, A. Omni-Dimensional Dynamic Convolution. *arXiv* **2022**, arXiv:2209.07947. [CrossRef]
36. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *arXiv* **2021**, arXiv:2103.14030. [CrossRef]
37. Pan, X.; Ge, C.; Lu, R.; Song, S.; Chen, G.; Huang, Z.; Huang, G. On the Integration of Self-Attention and Convolution. *arXiv* **2021**, arXiv:2111.14556.
38. Liu, S.; Huang, D.; Wang, Y. Learning Spatial Fusion for Single-Shot Object Detection. *arXiv* **2019**, arXiv:1911.09516. [CrossRef]
39. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression. *arXiv* **2019**, arXiv:1902.09630. [CrossRef]
40. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 12993–13000. [CrossRef]
41. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *IEEE Trans. Cybern.* **2021**, *52*, 8574–8586. [CrossRef]
42. Zhang, Y.-F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [CrossRef]
43. Gevorgyan, Z. SIoU Loss: More Powerful Learning for Bounding Box Regression. *arXiv* **2022**, arXiv:2205.12740. [CrossRef]
44. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* **2023**, arXiv:2301.10051.
45. Zhang, H.; Xu, C.; Zhang, S. Inner-IoU: More Effective Intersection over Union Loss with Auxiliary Bounding Box. *arXiv* **2023**, arXiv:2311.02877.
46. Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. SAR Ship Detection Dataset (SSDD): Official Release and Comprehensive Data Analysis. *Remote Sens.* **2021**, *13*, 3690. [CrossRef]
47. Wang, Y.; Wang, C.; Zhang, H.; Dong, Y.; Wei, S. A SAR Dataset of Ship Detection for Deep Learning under Complex Backgrounds. *Remote Sens.* **2019**, *11*, 765. [CrossRef]
48. Yu, C.; Shin, Y. An efficient YOLO for ship detection in SAR images via channel shuffled reparameterized convolution blocks and dynamic head. *ICT Express* **2024**, *10*, 673–679. [CrossRef]
49. Yasir, M.; Shanwei, L.; Mingming, X.; Jianhua, W.; Nazir, S.; Islam, Q.U.; Dang, K.B. SwinYOLOv7: Robust ship detection in complex synthetic aperture radar images. *Appl. Soft Comput.* **2024**, *160*, 111704. [CrossRef]
50. Chen, Z.; Liu, C.; Filaretov, V.; Yukhimets, D. Multi-Scale Ship Detection Algorithm Based on YOLOv7 for Complex Scene SAR Images. *Remote Sens.* **2023**, *15*, 2071. [CrossRef]
51. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *arXiv* **2016**, arXiv:1610.02391.