An Improved Ship Detection Algorithm Based on YOLOv8 for SAR Images

Xiaorui Wang*, Yuhan Wang, Xinwei Liu

School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou, China, 221116

*Corresponding author: 07224484@cumt.edu.cn

Abstract. Existing synthetic aperture radar (SAR) ship target detection algorithms are plagued by two primary issues: low detection accuracy and leakage. These issues stem from factors such as fuzzy target images, complex backgrounds, and a paucity of texture features of the target.To address these problems, this paper proposes a high-precision SAR ship target detection algorithm based on YOLOv8. The replacement of the C2f module in the YOLOv8 backbone with a CG-block module, which integrates local and global features, enhances the detection accuracy. This is due to the fact that the wide field of view of SAR images and the small size of ship targets necessitate an improvement in the detection accuracy.The neck part of the model is strengthened by the Gather-and-Distribute mechanism in the Gold-YOLO network, which improves the detection performance of small targets. The model incorporates the InnerSIoU loss function to enhance the regression accuracy, convergence speed, and model adaptability to complex scenes.Experimental results on the SSDD (Synthetic Aperture Radar Ship Detection Dataset) demonstrate that the enhanced algorithm attains a mean accuracy (mAP) of 98.42% and an accuracy of 96.21%, effectively achieving high-precision detection of ship targets in SAR images.

Keywords: Synthetic Aperture Radar, Ship Target Detection, Deep Learning, Feature Fusion, Loss Function.

1. Introduction

The sea area is vast and the climate is highly variable, often with extreme weather conditions. Synthetic aperture radar (SAR) can operate around the clock, offering high resolution and cloud penetration, making SAR imaging technology widely used in marine monitoring [1]. However, due to unique noise interference and the lack of distinct texture features in SAR images, ship targets often appear blurry and have low contrast [2]. Moreover, the complexity and variability of the marine background—such as waves, islands, tracks, and port infrastructure—further complicate ship detection [3]. In this context, enhancing the precision and dependability of vessel target recognition in synthetic aperture radar (SAR) images has emerged as a pivotal research challenge.

In recent years, deep learning technology has advanced rapidly, and target detection algorithms using convolutional neural networks (CNNs) have gradually become the focus of research on synthetic aperture radar (SAR) vessel identification. In particular, the YOLO series of algorithms have gained popularity in various target detection tasks due to its superior efficiency and real-time capabilities, showing strong adaptability and advantages in SAR image ship detection [4][5]. Several studies have applied classical YOLO models, such as YOLOv5, YOLOv7, and YOLOv8, to achieve better ship detection results in SAR images. For example, Sun and Wen (2024) introduced a streamlined ship detection algorithm leveraging YOLOv5 and GhostNet, thereby achieving a minimization of the parameter count without a concomitant compromise in accuracy [6]. Chen et al. (2024) proposed an improved version of the YOLOv7 model based on the coordinate attention mechanism and NWD metrics in the study of YOLOv7, which further enhances the detection of small targets [7]. Li et al. (2022) proposed the DFF-YOLOv5 algorithm for nearshore ship detection in SAR images, which improves detection performance in complex ocean backgrounds by using multi-scale feature fusion [8]. Subsequently, Sun et al. (2024) integrated edge enhancement and attention mechanisms into the YOLOv8 network, improving its ability to capture critical information in complex contexts and significantly reducing false and missed detections [9]. Zhao et al. (2024)

introduced RA-YOLO, an object detection algorithm that combines the receptive field attention mechanism with global information, enhancing model accuracy and generalization through this fusion [10].

However, existing models still suffer from limitations in detecting small targets with high accuracy and adaptability to complex backgrounds in SAR images. They are particularly affected by high-noise backgrounds, and the detection accuracy of multi-scale targets remains suboptimal. This study proposes a ship detection technique that utilizes an enhanced YOLOv8n model to overcome existing challenges. The objective of this approach is to enhance detection precision and improve the identification of small targets in SAR imagery. First, given the characteristics of SAR images, the C2f layer in the YOLOv8 model is replaced with the Context Guided Block (CG-Block) module [11]. The CG-Block module fuses local and global features, thereby enhancing the detection accuracy.Secondly, we introduce the Gather-and-Distribute mechanism (GD)[12] from the Gold-YOLO model to improve the Neck structure of YOLOv8, enhancing information fusion and enabling better performance in detecting small targets. Finally, the InnerSIoU loss function[13] is employed to improve the model's regression accuracy, convergence speed, and adaptability to complex scenarios. Experimental results on the SAR Ship Detection Dataset (SSDD) demonstrate that the improved algorithm achieves an average accuracy (mAP) of 98.42% and a detection accuracy of 96.21%, significantly outperforming the original YOLOv8 model and achieving high-precision ship detection in SAR images.

2. Improved algorithm design

2.1. Improved network architecture

The present study proposes an upgraded network architecture, as illustrated in Figure 1. Based on the original YOLOv8 architecture, the CG-block module [11] is introduced in the backbone to replace the original C2f module. The objective is to identify local characteristics, nearby context, and overarching contextual features, seamlessly integrating them to enhance the precision of object detection. Gold-YOLO is a sophisticated object detection framework that utilizes an innovative GD mechanism to improve the efficiency of data fusion. [12]. Through convolution and self-attention operations, the mechanism processes feature information from different layers of the network, achieving effective fusion of multi-scale features. This results in high accuracy while maintaining low latency. Additionally, Gold-YOLO introduces MAE-style pre-training for the first time in the YOLO series, further improving the model's learning efficiency and accuracy. In the Neck module, the original Neck of YOLOv8 is replaced with the Gold-YOLO Neck, and the GD mechanism is employed to fuse features at different levels, enabling efficient information interaction. This strengthens the model's ability to fuse information in the Neck, thereby improving small target detection performance. Finally, the fused InnerSIoU loss function is used for training to enhance the model's convergence speed and regression accuracy [13].

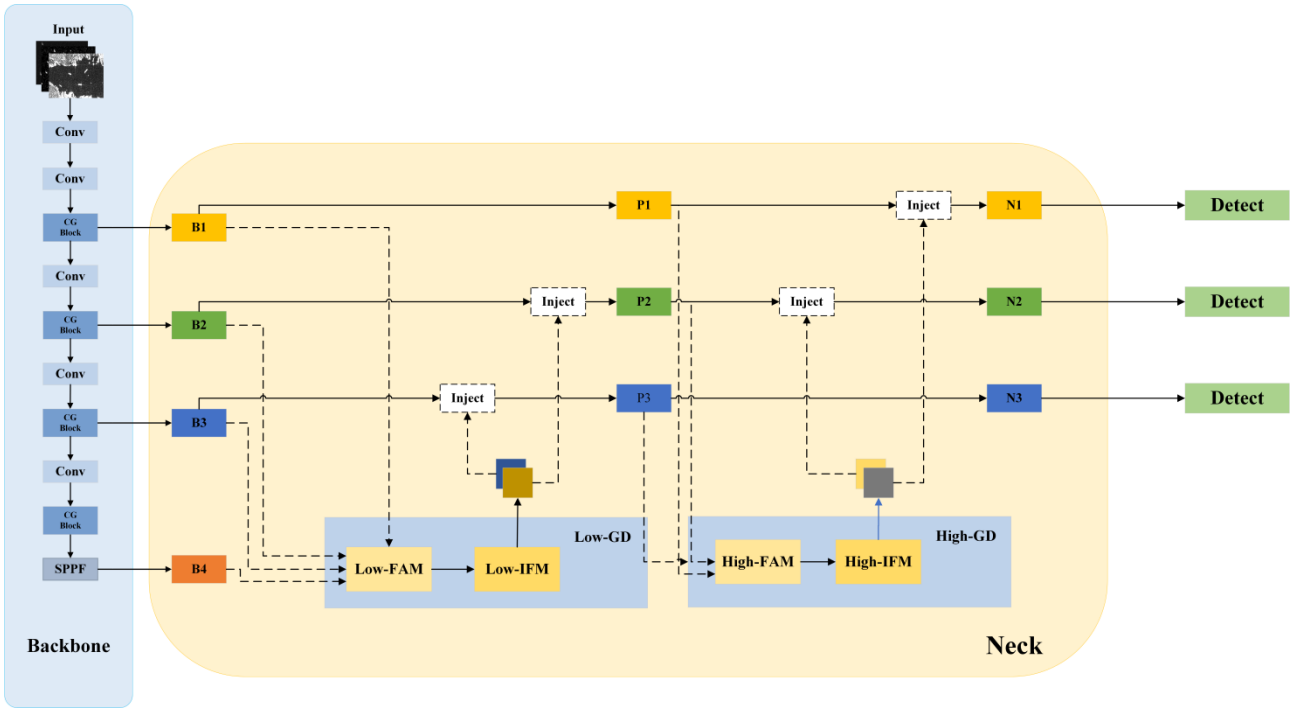


Fig 1. Improved algorithm structure in this paper

2.2. CG Block

CG Block is used to capture the local features, surrounding context features, and global features of the input image, and fuse these three different levels of semantic features to improve object detection accuracy. This module consists of the following four parts:(1) Local Feature Extractor (F_{loc}): Extracts local features using standard convolutional layers. (2) Peripheral Context Extractor (F_{sur}): Uses an expanded convolutional layer to obtain a larger receptive field for extracting surrounding context features. (3) Joint Feature Extractor (F_{joi}): The features extracted by Floc and Fsur are concatenated, and features fused with the surrounding context are obtained through batch normalization (BN) and the PreLU activation function. (4) Global Context Extractor (F_{glo}): Global context features are fused through the Global Average Pooling Layer (GAP), and the complex nonlinear relationships between features are further learned through two Fully Connected Layers (FC) to refine the global context features.

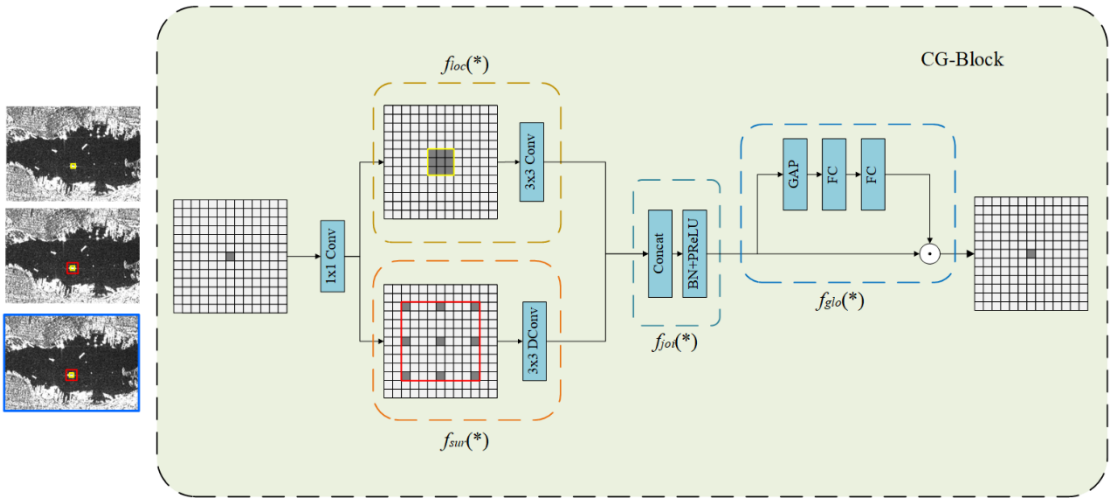


Fig 2. CG Block structure

Through these components, CGBlock can efficiently capture features at various scales in complex scenes, enabling the improved network to perform more accurate object detection. By introducing the

CGBlock module into the Backbone of the YOLOv8 network, the model can establish connections between local and global context features, thereby improving the accuracy of ship detection in SAR images.

2.3. Gather-and-Distribute mechanism (GD)

GD is a key component of the Gold-YOLO model, primarily designed to efficiently address the problem of information fusion [11]. In this mechanism, the Feature Alignment Module (FAM) and Information Fusion Module (IFM) are introduced to effectively aggregate features from different levels. FAM aligns features to ensure that features from different layers can be compared and fused in the same dimension, while IFM integrates these aligned features, creating a richer and more meaningful feature representation. The fused information is then distributed back to all levels of the network through the Inject Module. This process ensures a smoother flow of information between features at each level, enabling the effective combination of low-level local information and high-level global context. As a result, the Gold-YOLO model can make more comprehensive use of multi-scale features, significantly improving object detection accuracy while maintaining low latency. This mechanism not only enhances the model's ability to detect small targets but also improves overall performance, highlighting the critical role of the GD mechanism in target detection. Below is a detailed description of the three key modules in the Gold-YOLO architecture.

2.3.1 Low-GD

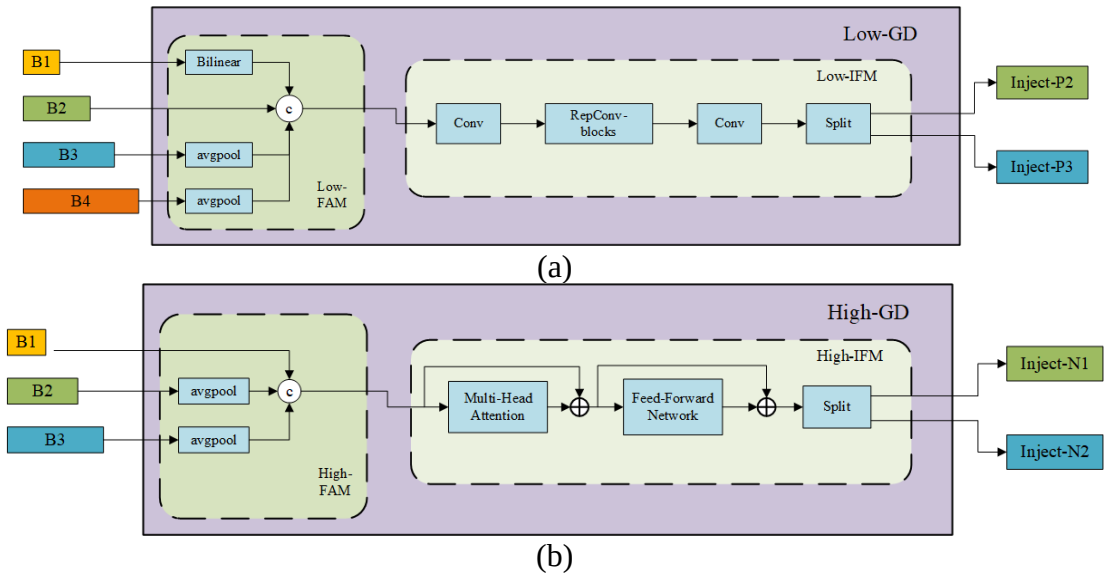


Fig 3. GD module structure

In this particular branch, the output features of the backbone network, designated as B1, B2, B3, and B4, are integrated to generate high-resolution features that maintain the integrity of small target information. The structural configuration is illustrated in Figure 3(a).

Low-Stage Feature Alignment Module (Low-FAM): The Low-Stage Feature Alignment Module (Low-FAM) uses the AvgPool operation to downsample the input features, ensuring feature size consistency. The minimum feature size ($R_{B2} = \frac{1}{4}R$) is determined by adjusting the features to align with the cluster, This adjustment results in the attainment of the aligned feature Fa. The goal of the Low-FAM technique is to efficiently integrate information while minimizing computational complexity, achieving a balance between speed and accuracy.

Low-stage Information Fusion Module (Low-IFM): This module combines multi-layer reparameterized convolutional blocks (RepBlocks) and separation operations to achieve efficient information fusion. Specifically, RepBlock takes the aligned feature F_a (number of channels = $C_{B1} + C_{B2} + C_{B3} + C_{B4}$) as input and generates the fusion feature F_f (number of channels = $C_{B1} + C_{B2}$).

The features generated by RepBlock are then split along the channel dimension into $F_{inj-P3,}$ and F_{inj-P4}, which are subsequently integrated with features from disparate levels.

The formula is as follows:

$$F_a = Low\_FAM([B1, B2, B3, B4]), \tag{1}$$

$$F_f = RepBlock(F_a), \tag{2}$$

$$F_{inj-P3,}F_{inj-P4} = Split(F_f). \tag{3}$$

Through this design, Low-IFM effectively integrates multi-level information, enhancing the expressiveness and performance of the model.

2.3.2 High-GD

As illustrated in Figure 3(b), High-GD integrates the features $\{P_1, P_2, P_3\}$ that are generated by Low-GD.

High-Feature Alignment Module (High-FAM): Uniform feature size is achieved by reducing the dimensionality of input features using average pooling. In particular, when the input feature size is $\{R_{P1}, R_{P2}, R_{P3}\}$, average pooling reduces the feature size to the smallest size within the feature group ($R_{P1} = \frac{1}{8}R$). By integrating information through feature pooling, this process not only effectively reduces redundancy in the input features but also decreases the computational burden in the subsequent steps of the Transformer module.

High-Stage Information Fusion Module (High-IFM): The High-IFM consists of a Transformer block and a segmentation operation, which is divided into three steps: (1) F_f is obtained by fusing the alignment feature F_a, output by the High-FAM module, through the Transformer block.(2) A 1×1 convolution operation is applied to compress the number of channels of F_f to Sum(C_{P1},C_{P2}). (3) F_f is then split into F_{inj_N1} and F_{inj_N2} along the channel dimension using the segmentation operation, and is further integrated with the features from the current level.

The formula is as follows:

$$F_a = High\_FAM([P1, P2, P3]), \tag{4}$$

$$F_f = Transformer(F_a), \tag{5}$$

$$F_{inj\_N1}, F_{inj\_N2} = Split\left(Conv1 \times 1(F_f)\right). \tag{6}$$

2.3.3 Information injection module (Inject)

This module employs an attention mechanism to fuse information, thereby more efficiently integrating global data at every level, as depicted in Figure 4. Specifically, the inputs include local information (features from the current hierarchy) and global injection information (generated by the IFM), denoted as F_{local} and F_{inj} respectively. Two different convolution operations are applied to F_{inj} to obtain F_{global_embed} and F_{act}, while F_{local} is processed through convolution to produce F_{local_embed}. The attention mechanism is then used to fuse these features, generating F_{out}.

Since the dimensions of F_{local} and F_{global} differ, average pooling or bilinear interpolation is used to adjust the sizes of F_{global_embed} and F_{act} ensuring that they are aligned with F_{inj} for correct matching.Subsequent to each attention fusion step, RepBlock is incorporated to extract and integrate information. (Note: The information injection module in High-GD is identical to that in Low-GD. At the high stage,F_{local} is equal to P_i)

Below is the formula:

$$F_{global\_act\_Pi} = resize\left(Sigmoid\left(Conv_{act}\left(F_{inj\_Pi}\right)\right)\right) \tag{7}$$

$$F_{global\_embed\_Pi} = resize\left(Conv_{global\_embed\_Pi}\left(F_{inj\_Pi}\right)\right) \tag{8}$$

$$F_{att\_fuse\_Pi} = Conv_{local\_embed\_Pi}(Bi) * F_{ing\_act\_Pi} + F_{global\_embed\_Pi} \quad (9)$$

$$Pi = RepBlock(F_{att\_fuse\_Pi}) \quad (10)$$

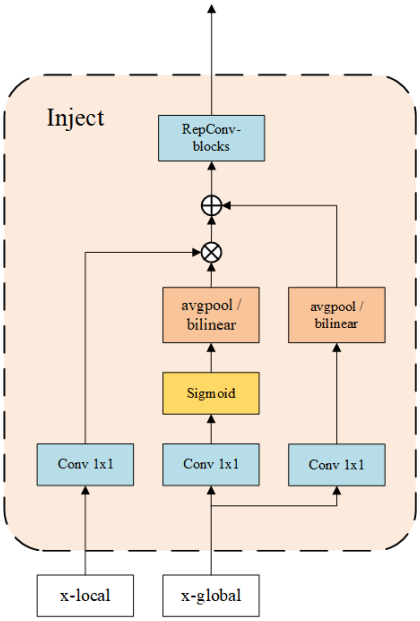


Fig 4. Inject module structure

2.4. InnerSIoU Loss Function

Within the YOLOv8 framework, the CIoU loss function (Complete Intersection over Union) quantifies the discrepancy between predicted and actual bounding boxes by incorporating IoU, center point distances, and aspect ratio discrepancies. The formula is:

$$CIoU = IoU - \frac{\rho^2(b,b_{gt})}{c^2} - \alpha \cdot v \quad (11)$$

In this context, $\rho(b\ ,\ b_{gt})$ is defined as the Euclidean distance separating the centers of the predicted box and the ground truth box, and c is the diagonal length of their smallest circumscribed rectangle, v measures the aspect ratio difference and α is used to adjust the influence of the aspect ratio.

When the aspect ratio of the predicted box is close to that of the ground truth box, v approaches 0, which weakens the penalty effect of CIoU and, in turn, slows down the model's convergence. Therefore, the CIoU loss function is not well-suited for ship detection tasks in SAR images dominated by small targets.

InnerSIoU is an improved loss function designed to optimize the matching degree in bounding box prediction, particularly for the internal geometric alignment between the predicted and ground truth boxes [12]. Unlike traditional IoU, DIoU and CIoU, InnerSIoU enhances the accuracy of bounding box regression by incorporating IoU, center point distance, and edge alignment terms.

The formula is as follows:

$$InnerSIoU = IoU - \frac{\rho^2(b,b_{gt})}{c^2} - \beta \cdot EdgeAlign - \alpha \cdot v \quad (12)$$

IoU represents the ratio of the intersecting area to the union area of the predicted and ground truth bounding boxes.$\rho(b\ ,\ bgt)$ is the Euclidean distance between the center points of the predicted and ground truth boxes, used to measure the position error. c is the diagonal length of the smallest circumscribed rectangle containing both the predicted and ground truth boxes. $EdgeAlign$ is a boundary alignment term that evaluates the degree of overlap between the edges of the predicted box and the ground truth box, ensuring better alignment with the actual target. v measures the aspect

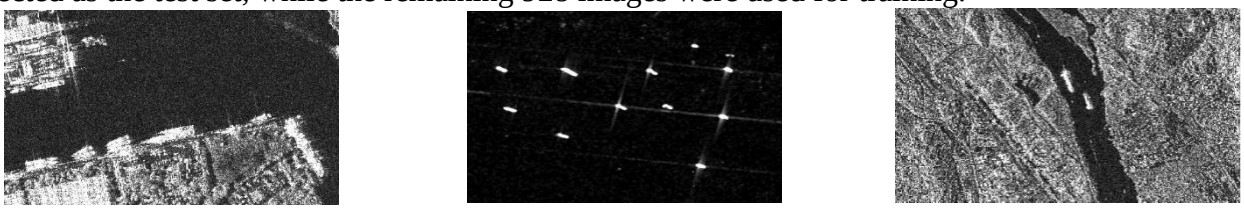
ratio difference between the predicted and ground truth boxes. α and β are weight factors that adjust the impact of the aspect ratio and edge alignment terms.

With this combination, InnerSIoU effectively constrains center offset, edge alignment, and aspect ratio differences, while preserving the overlapping area. This enhances both the positioning and shape matching accuracy of the bounding box. This design is particularly well-suited for ship target detection in SAR images, where small targets are prevalent.

3. Results

3.1. Experimental data and Setup

In this paper, the public dataset SSDD is used to evaluate the effectiveness of the network. The dataset consists of 1,160 images from three different SAR satellites, with a total of 2,587 ships, averaging 2.23 ships per image. The SSDD dataset includes images with different polarization modes, resolutions, locations, and scenes, featuring ship targets of varying sizes, as shown in Figure 5, which provides a typical example. During the experiment, 232 images, with mantissa numbers 1 and 9, were selected as the test set, while the remaining 928 images were used for training.



(a) Ship targets in a complex context (b) Dense small-sized ship targets (c) Ship targets under heavy noise

Fig 5. shows a typical example of three different scenarios

Table.1. Experimental environment and experimental parameters

Item	Content
operating system	Windows 11
Development environment	Python 3.9 、Pytorch 1.13.0 、CUDA 12.0
Graphics Card (GPU)	NVIDIA GeForce GTX 3060 , 12Gvideo memory
epochs	200
Batch-size	16

3.2. Experimental evaluation metrics

The present study utilizes precision (P), recall (R), mean average precision (mAP), and the number of parameters (Params) as the metrics for evaluation. In this context, TP and FP represent the number of true positive and false positive instances, respectively, while FN denotes the count of false negatives. The following equations are used to calculate precision P, recall R, and mAP:

$$precision = \frac{TP}{TP+FP} \tag{13}$$

$$Recall = \frac{TP}{TP+FN} \tag{14}$$

$$AP = \int_0^1 P(R)dR \tag{15}$$

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i \tag{16}$$

In the above formulas, TP (True Positive) represents the number of correctly detected positive samples, while FP (False Positive) refers to the number of samples incorrectly detected as positive. FN (False Negative) denotes the number of positive samples that were missed. C is the total number of detected categories, and AP_i represents the average precision of the i-th category.

3.3. Ablation experiments

To verify the effectiveness of the CG-Block module, the Gold-YOLO (GD) module, and the InnerSIoU loss function, four sets of ablation experiments were conducted on the SSDD dataset. The experimental results are shown in Table 2. Experiment 1 presents the detection results of the original YOLOv8 network, while the remaining experiments correspond to the detection results using the aforementioned improved modules. A "√" indicates that the relevant method was applied, and the bolded values in the table represent the best results.

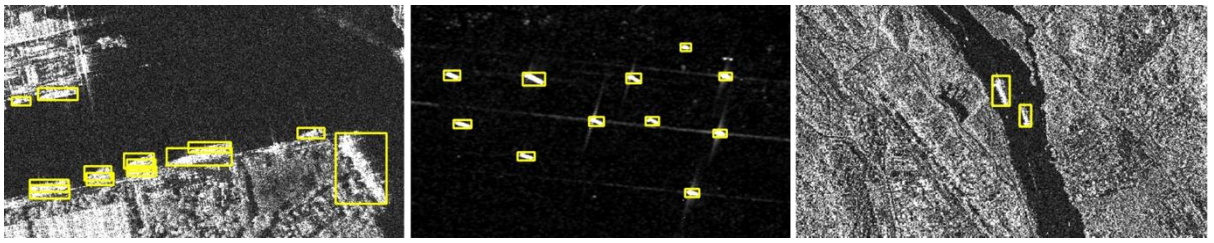
Table.2. Results of ablation experiments for CG-Block, GD, and InnerSIoU

experiment	CG-Block	GD	InnerSIoU	P(%)	R(%)	mAP(%)	Parameter
1	×	×	×	95.71	89.99	96.96	3005843
2	√	×	×	92.61	**93.41**	97.45	**2583024**
3	√	√	×	96.09	**93.41**	97.88	7637744
4	√	√	√	**96.21**	92.92	**98.42**	7637744

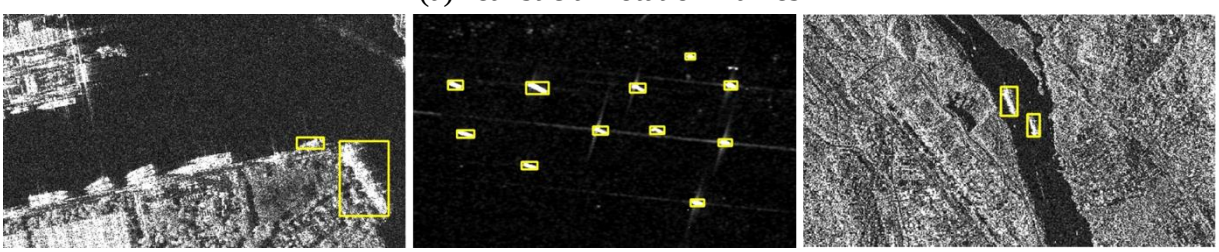
It can be observed that with the introduction of the CG-Block, GD module, and the InnerSIoU loss function, both P and mAP improve significantly. Specifically, P increases from 95.71% to 96.21%, and mAP rises from 96.96% to 98.42%, and the number of parameters increases within an acceptable range. The detection accuracy of the improved algorithm proposed in this paper on the SSDD dataset is significantly better than that of the original YOLOv8 model.

3.4. Verification of prediction results

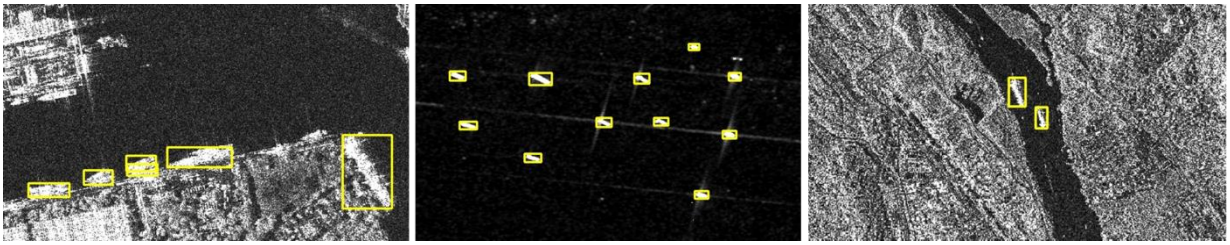
To more intuitively compare the ship target detection performance of the proposed method with the original YOLOv8 network, Figure 6 shows the ground truth bounding boxes in three different scenarios, along with the detection results of both the YOLOv8 network and the proposed method. In the complex background of docking detection (column 1), the original YOLOv8 network struggles to accurately distinguish between ships and docking buildings due to blurred image details and textures, leading to both missed and false detections. In contrast, with the introduction of the CG-Block and GD modules, the proposed method enhances the feature extraction capabilities of the network in complex backgrounds, effectively reducing the occurrence of missed and false detections for ship targets. In other scenarios (columns 2 and 3), the detection bounding boxes of the proposed method are more accurate in both position and size compared to the original YOLOv8 network, leading to improved detection accuracy.



(a)Realistic annotation frames



(b)YOLOv8



(c)Proposed method

Fig 6. Comparison of the results of YOLOv8 and proposed method in three scenarios

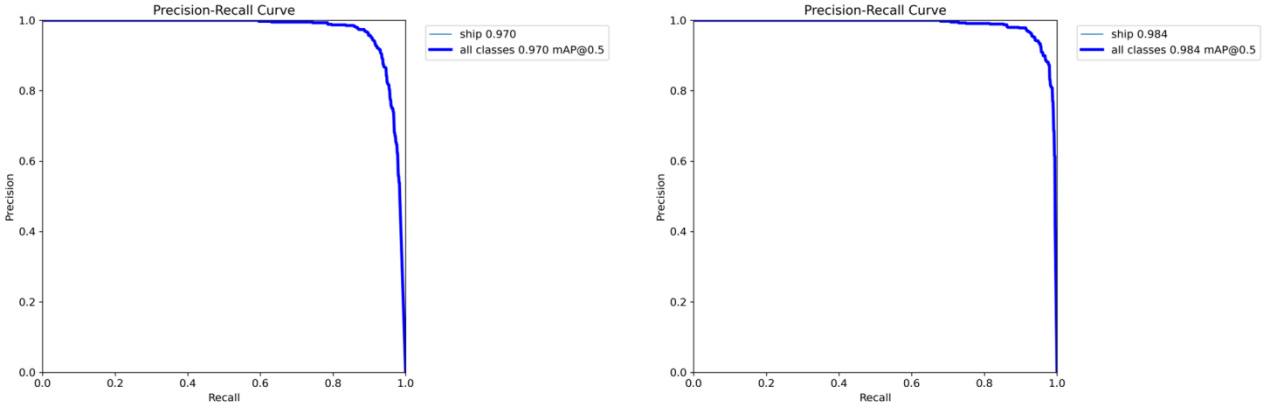


Fig 7. Precision-Recall curve comparison: YOLOv8 vs. proposed method

In summary, compared with the original YOLOv8 network, the proposed method demonstrates excellent performance in addressing issues such as image blurring and insufficient texture features in SAR image ship target detection. It also performs exceptionally well across different resolutions and scenes.

4. Conclusion

This paper addresses the challenges in ship target detection from SAR images, including image blur, noise, the presence of numerous dense small targets in offshore areas, and the mixed clarity of targets and buildings in complex coastal backgrounds, all of which contribute to missed detections and suboptimal accuracy. We propose a ship target detection method based on an improved YOLOv8 model for SAR images. A series of ablation experiments are conducted. using the publicly available SSDD dataset. The experimental results demonstrate that the addition of the CG-Block module, the information GD from Gold-YOLO, and the InnerSIoU loss function all significantly improve the accuracy of ship detection in SAR images. Specifically, the mAP increases to 97.45%, 97.88%, and 98.42%, respectively.

Moreover, compared to the original YOLOv8 network, the improved algorithm achieves the mAP of 98.42% and a P of 96.21%, surpassing the original model by 1.46 and 0.5 percentage points, respectively. These results confirm the superiority of our method in detecting targets under conditions of high noise, varying resolutions, different scenes, and ship targets of varying sizes.

References

[1] Zhang T, Zhang X, Ke X, et al. HOG-ShipCLSNet: A novel deep learning network with hog feature fusion for SAR ship classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-22.

[2] Wang Haoyu, Yang Haitao, Wang Jinyu, et al. A review of denoising methods for remote sensing images[J]. Computer Engineering and Applications, 2024, 60(15):55-65.

[3] Dai W, Mao Y, Yuan R, et al. A novel detector based on convolution neural networks for multiscale SAR ship detection in complex background[J]. Sensors, 2020, 20(9): 2547.

[4] Li Bo, Li Zhikang, Zhou Yubin. SAR Ship Detection Algorithm Based on Feature Fusion and Attention Mechanism[J]. Electronic Measurement Technology, 2024, 47(10):134-140.

[5] Jiang Fukun, Huang Xiangcheng, Zhang Xiaobo, et al. Research on Ship Target Detection Method for SAR Based on YOLO Model. Journal of Ocean Technology, 2023, 42(04): 14-27.

[6] Sun Peishuang, Wen Xianbin. Ship Target Detection Algorithm for SAR Images Based on Improved YOLOv5 Model. Electro-Optics and Control, 2024, 31(08): 32-37+85.

[7] Chen Wenhan, Zhu Zhengwei, Song Changlong. Ship Target Detection Method for SAR Images Based on Improved YOLOv7. Electro-Optics and Control, 1-11 [2024-11-01].

[8] Li Yonggang, Zhu Weigang, Huang Qiongnan, et al. Ship Target Detection in Nearshore SAR Images with Complex Backgrounds. Systems Engineering and Electronics Technology, 2022, 44(10): 3096-3103.

[9] Sun Shanshan, Zhang Lijuan, Zhao Hui. SAR Ship Detection Model Based on Edge Enhancement and Attention Mechanism. Electro-Optics and Control, 2024, 31(08): 92-97+110.

[10] Zhao Jingyu, Li Min, Chen Xiefa, et al. SAR Ship Target Detection Algorithm Integrating Attention Mechanism and Global Information. Journal of Rocket Force Engineering University, 2024, 38(04): 40-46.

[11] Wu T, Tang S, Zhang R, et al. Cgnet: A light-weight context guided network for semantic segmentation[J]. IEEE Transactions on Image Processing, 2020, 30: 1169-1179.

[12] Wang C, He W, Nie Y, et al. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism[J]. Advances in Neural Information Processing Systems, 2024, 36.

[13] Zhang H, Xu C, Zhang S. Inner-IoU: more effective intersection over union loss with auxiliary bounding box[J]. arXiv preprint arXiv:2311.02877, 2023.