Global Context-enhanced YOLO11 for Ship Detection in SAR Imagery

Chao Ma

School of Information Technology Engineering, Tianjin University of Technology and Education, Tianjin 300222, China

m18706181119@163.com

Abstract

Ship detection in Synthetic Aperture Radar (SAR) images holds significant practical value in fields such as maritime security, military reconnaissance, and commercial shipping monitoring. However, the inherent speckle noise in SAR images, complex background interference, and the diverse scales and varying orientations of ship targets pose severe challenges to efficient and accurate ship detection. Traditional deep learning models, particularly the YOLO series of detectors, often struggle to fully capture global contextual information when processing these complex scenes, thereby compromising detection performance—especially regarding small and densely clustered targets. To address this issue, this study proposes an enhanced SAR ship detection method centered on the integration of an optimally designed global context attention module. This module is designed to effectively capture global contextual information at a low computational cost and fuse it back into the feature maps. By introducing this optimized module into the YOLO11 detector, the model significantly improves the feature discrimination of ship targets in SAR images and its robustness in complex backgrounds. Experimental results demonstrate that this method effectively enhances the accuracy and generalization capability of SAR ship detection while maintaining computational efficiency.

Keywords

YOLO11, SAR Ship Detection, Global Context Attention, Deep Learning.

1. Introduction

Synthetic Aperture Radar (SAR) technology, with its all-weather, all-time imaging capabilities, has become a key technology in fields such as marine environmental monitoring, maritime traffic management, fisheries surveillance, and military reconnaissance[1,2,3,4,5]. Compared to optical sensors, SAR can penetrate clouds, haze, and darkness to provide continuous and reliable maritime information, which holds irreplaceable strategic value for the real-time monitoring and identification of vessels. However, the unique imaging mechanism of SAR imagery also presents a series of inherent challenges, significantly limiting its performance in ship detection missions.

First, the coherent speckle noise commonly found in SAR images is a multiplicative noise that degrades image quality, blurs edges, and causes loss of detail, severely hindering the effective feature extraction of ship targets[6,7]. Second, ship targets in SAR images typically exhibit extreme multiscale characteristics; ranging from small fishing boats to large aircraft carriers, the differences in their pixel sizes within the image are enormous, making it difficult for detection models to capture features at different scales[8,9]. Furthermore, the complex marine clutter background, confusion with coastal features and land objects, and variations in ship scattering characteristics under different attitude angles all contribute to high false positive and false negative rates in SAR ship detection[10,11].

In recent years, deep learning—particularly convolutional neural networks—has achieved breakthrough progress in the field of object detection. The You Only Look Once (YOLO) series of algorithms has become the mainstream paradigm for real-time object detection tasks due to its exceptional balance of speed and accuracy. From early versions of YOLO to the latest YOLOv11, this series of models has continuously improved detection performance and end-to-end deployment capabilities by introducing more efficient architectures, optimization strategies, and data augmentation techniques. As the latest iteration in the YOLO series, YOLOv11 has demonstrated state-of-the-art (SOTA) performance across multiple benchmarks and shows great promise in various application domains, including power equipment detection[12].

Although the YOLO series of models performs exceptionally well in optical image object detection, applying them directly to ship detection in SAR images still poses numerous challenges. Traditional local receptive field designs make it difficult for these models to effectively capture the long-range dependencies and global contextual information that are critical in SAR images[4]. For example, in port or complex coastline scenarios, the sparsity of ship targets, their dense arrangement, or their weak contrast with the background all require the model to possess stronger global awareness for accurate classification. Existing research has attempted to enhance model performance by introducing attention mechanisms, such as channel attention or spatial attention modules[1]. However, these mechanisms often operate in a local or decoupled manner, making it difficult to fully model global semantic relationships. This limits the model's ability to comprehensively understand and distinguish ship features in complex SAR scenarios[5].

Given these challenges, this study aims to deeply integrate a global context-aware attention mechanism into the YOLOv11 architecture to build a ship detection framework that exhibits high robustness and accuracy under complex sea conditions. Our key contributions are as follows:

(a)Optimization and Integration of the Global Context Attention Module: We propose an innovative global context attention module that employs a dual-mask design to separately learn "feature compensation" and "feature filtering" mechanisms. This enables the module to efficiently capture and integrate global contextual information with low computational overhead, thereby enhancing the model's ability to represent the features of ship targets[13].

(b)Adaptive Improvements to the YOLOv11 Architecture: The optimized GC module is seamlessly integrated into the backbone network and feature fusion layer of YOLOv11 to fully leverage the advanced architectural advantages of YOLOv11 and enhance its robustness against challenges specific to SAR imagery.

Through these improvements, this study aims to significantly enhance the accuracy, robustness, and efficiency of ship detection in SAR imagery, thereby providing more reliable technical support for intelligent marine monitoring and offering new insights for future research on target detection based on SAR imagery.

2. Materials and Methods

2.1. Improving the Overall Architecture of the Model

To address the challenges posed by severe coherent speckle noise affecting ship targets in Synthetic Aperture Radar images, complex background clutter (such as waves and port structures), and a wide range of target sizes, this paper proposes an enhanced ship detection architecture based on YOLO11.

As the current state-of-the-art single-stage detector, YOLO11 employs an efficient C3k2 module and SPPF architecture in its backbone. However, due to the limited local receptive field of convolutional neural networks, this model lacks the ability to model global spatial context when processing isolated bright targets against a wide-area maritime background. To address this, this paper introduces a dual-path global context attention module (DGC-Block) at the output of the backbone. The improved architecture effectively captures long-range feature dependencies, achieving clutter suppression and calibration of ship features.

2.2. Enhanced Dual-Path Global Context Module

2.2.1. Design Motivation

In SAR images, the scattering characteristics of ship targets are significantly influenced by the angle of incidence and material composition, often appearing as irregular bright regions. Although traditional non-local neural networks can establish global dependencies, their computational complexity increases quadratically with the size of the feature map. The DGC-Block introduced in this paper combines the global modeling capabilities of Non-local Networks with the efficiency of Squeeze-and-Excitation (SE) networks, and further innovatively designs a dual-path mechanism comprising additive compensation and multiplicative filtering.

2.2.2. Mathematical Modeling

Let the input feature map be $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$, where C denotes the number of channels, and H and W denote the height and width of the feature map, respectively. To model global dependencies without significantly increasing the computational burden, the Dual-path Global Context Block (DGC-Block) designed in this paper consists of three main stages: global context modeling, bottleneck feature transformation, and composite gated fusion.

The global context modeling stage aims to compress the spatial information of the entire image into a global descriptor. Unlike traditional global average pooling, this paper adopts an attention-based pooling approach that focuses on key scattering regions by learning spatial weight distributions. The module concurrently constructs weights for the additive path w_{add} and the multiplicative path w_{mul}:

$$\alpha_j^{add} = \frac{\exp(\mathbf{W}_{mask}^{add} x_j)}{\sum_{m=1}^{H\times W} \exp(\mathbf{W}_{mask}^{add} x_m)} \quad (1)$$

$$\alpha_j^{mul} = \frac{\exp(\mathbf{W}_{mask}^{mul} x_j)}{\sum_{m=1}^{H\times W} \exp(\mathbf{W}_{mask}^{mul} x_m)} \quad (2)$$

In particular, $\mathbf{W}_{mask}$ is a 1×1 convolutional kernel, and x_j represents the feature vector at the j spatial location in the feature map. By performing a weighted sum over all pixels, we obtain the corresponding global context feature vectors $\mathbf{G}_{add}$ and $\mathbf{G}_{mul}$:

$$\mathbf{G}_{add} = \sum_{j=1}^{H\times W} \alpha_j^{add} x_j, \quad \mathbf{G}_{mul} = \sum_{j=1}^{H\times W} \alpha_j^{mul} x_j \quad (3)$$

To capture nonlinear dependencies between channels and reduce computational complexity, the global vector $\mathbf{G}$ is fed into a bottleneck structure for transformation. This structure consists of two layers of 1×1 convolutions, with Layer Normalization (LN) introduced in between to stabilize the gradient distribution in deep networks. The transformation process is formalized as follows:

$$\psi(\mathbf{G}) = \mathbf{W}_{v2}\left(\mathrm{ReLU}\left(\mathrm{LN}\left(\mathbf{W}_{v1}\mathbf{G}\right)\right)\right) \quad (4)$$

In particular, $\mathbf{W}_{v1} \in \mathbb{R}^{\frac{C}{r}\times C}$ is used for channel dimension reduction, $\mathbf{W}_{v2} \in \mathbb{R}^{C\times\frac{C}{r}}$ is used for channel dimension expansion, and r is the reduction ratio (set to 2 in this paper). Compared to Batch Normalization, LN can normalize neurons within a single sample, making it more suitable for data sensitive to global statistical properties, such as SAR images.

This paper innovatively combines two fusion strategies to achieve refined refinement of feature maps:

$$\mathbf{Y} = \mathbf{X} \otimes \sigma(\psi_{mul}(\mathbf{G}_{mul})) + \psi_{add}(\mathbf{G}_{add}) \tag{5}$$

In this formula, $\otimes$ denotes element-wise multiplication across channels, and σ is the sigmoid activation function. The multiplicative path ($\mathbf{X} \otimes \sigma(\cdot)$) applies gated weighting to the original features using global information, serving as a “feature filter” to suppress background noise; the additive path ($+\psi(\cdot)$) acts as a “feature compensator,” enhancing the semantic strength of the target itself.

2.2.3. Mechanism Analysis

The dual-path design of DGC-Block provides solutions to two typical challenges in SAR ship detection:

Suppressing strong reflections from the sea surface: In SAR images, waves or islands often appear as bright clutter. The multiplicative path utilizes global contrast information to identify that these regions lack the scattering topology characteristic of ships, thereby reducing their weight and enhancing the detector’s robustness against interference.

Enhancing robustness for weak targets: For small vessels in offshore areas, their pixel coverage is extremely low, making them prone to being lost in deep convolutional layers. The additive path compensates for spatial context information in local feature maps by propagating global semantic features, thereby enhancing the model’s detection threshold for weak targets.

2.2.4. Integration and Superiority

In this paper, DGC-Block is integrated into the final layer of the YOLO11 backbone network. The advantages of this design are reflected in the following three aspects:

Semantic-level alignment: The backbone network extracts highly abstract semantic features at its output. Introducing global context at this stage enables the model to obtain the most comprehensive macro-level perspective before making object classification and localization decisions, which helps distinguish between land-based structures and ship targets that have similar shapes.

Controllability of computational complexity: As the network depth increases, the spatial resolution of feature maps gradually decreases. Performing global operations at the final layer involves the minimum number of spatial points ($H \times W$), which allows the DGC-Block to significantly improve accuracy while adding only a small number of parameters, ensuring real-time performance for end-to-end inference.

Smooth Feature Transition: Feature maps refined by DGC-Block possess a higher signal-to-noise ratio. When these are then fed into the Neck component for multi-scale feature fusion, this effectively prevents false alarms caused by the mixing of shallow-layer noise with deep-layer semantics.

3. Experiments and Results

3.1. Experimental Setup and Dataset Description

3.1.1. About the Dataset

This paper uses the publicly available SAR-Ship-Dataset for model validation. The dataset contains 43,819 ship tiles acquired by the GF-3 and Sentinel-1 satellites, with resolutions ranging from 3 m, 5 m, 8 m, 10 m, to 25 m. Characterized by complex backgrounds (including ports, nearshore areas, and the open sea), a wide range of target scales, and diverse imaging noise, this dataset enables a comprehensive evaluation of the robustness of detection algorithms in real-world SAR scenarios. In the experiments, the dataset was randomly split into training and testing and val sets in an 7:2:1 ratio. To ensure training efficiency, we randomly selected 4,000 images from the entire dataset for training and testing.

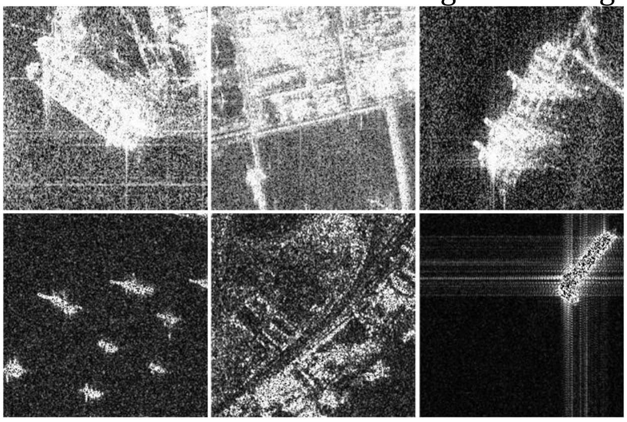

Fig. 1 SAR-Ship-Dataset

3.1.2. Test Environment and Configuration Settings

This experiment uses the Ultralytics deep learning framework to improve and validate algorithms. The experimental hardware environment consists of workstations equipped with NVIDIA GPUs, which utilize CUDA cores for computational acceleration. The software environment runs on the Windows operating system, and Python version 3.8 or higher is used for programming. The core deep learning libraries include PyTorch (>= 1.8.0) and Torchvision (>= 0.9.0). Additionally, OpenCV is used for image preprocessing, while Matplotlib and Seaborn are employed for the visualization and analysis of experimental results.

All input images are uniformly resized to 640 × 640 pixels before being fed into the network to balance detection accuracy and computational efficiency. To enhance the model's generalization ability and prevent overfitting, various data augmentation strategies were applied in the experiments, including mosaic augmentation (probability of 1.0), horizontal flipping (probability of 0.5), random translation (coefficient of 0.1), and random scaling (coefficient of 0.5).

The model is trained using the SGD (Stochastic Gradient Descent) optimizer, with an initial learning rate set to 0.01, combined with a cosine annealing strategy to dynamically adjust the learning rate, with the final learning rate factor set to 0.01. The momentum parameter is set to 0.937, and the weight decay coefficient is 0.0005. The total number of training epochs is set to 50, and the batch size is 4. To stabilize gradients during the early training phase, the first 3 epochs are set as the warm-up phase. The loss function consists of the bounding box regression loss, the classification loss, and distribution-focused loss, with weight coefficients set to 7.5, 0.5, and 1.5, respectively. The experiments enabled Auto-Mixing Precision (AMP) training to improve training efficiency and optimize GPU memory usage without sacrificing accuracy.

3.2. Evaluation Criteria

To objectively evaluate the detection performance of the improved model for SAR ship targets, this paper selects precision (P), recall (R), and mean average precision (mAP) as evaluation metrics. Among these, mAP_{50} represents the mean average precision when the intersection-over-union (IoU) threshold is set to 0.5, and its formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

$$mAP = \int_0^1 P(R)dR \tag{8}$$

In this context, TP, FP and FN represent true positives, false positives, and false negatives, respectively.

3.3. Quantitative Analysis

We conducted comparative experiments between the improved YOLO11 model described in this paper and the original YOLO11 baseline model on the SAR-Ship-Dataset. The specific experimental results are shown in Table 1.

Table 1 Object Detection Results

Model	P (%)	R (%)	mAP_{50} (%)	mAP_{50-95} (%)
Baseline	0.893	0.844	0.919	0.501
Ours	0.890	0.857	0.926	0.516

After introducing the dual-path global context attention module, the model's mAP_{50} improved by approximately 0.7%. This indicates that DGC-Block effectively enhances the model's ability to detect scattered features of ships by establishing dependencies among global pixels.

The improved model enhances recall while maintaining high accuracy. This indicates that global contextual information helps the model distinguish weak targets from complex backgrounds—such as strong marine clutter and land structures—thereby reducing false negatives.

4. Conclusion

This paper addresses the challenges of complex background interference and missed detection of small targets in ship detection using Synthetic Aperture Radar (SAR) images by researching and implementing a detection framework based on an improved YOLO11 model. By introducing a global context attention mechanism into the model's backbone network, this paper constructs a feature extraction scheme capable of effectively capturing long-range dependencies, thereby addressing the limitations of traditional convolutional operations in global modeling. The research focuses on designing a dual-path fusion module with additive compensation and multiplicative filtering functions, which effectively suppresses sea clutter and enhances the semantic information of weak and small targets through feature rectification. Experimental results on public datasets demonstrate that the improved model performs excellently, with significant improvements in all core detection metrics compared to the baseline model, fully proving the critical role of global contextual information in enhancing the robustness of target detection in SAR images. The findings of this study not only provide an efficient and accurate technical solution for maritime surveillance in complex environments but also validate the superiority of combining lightweight detection models with enhanced attention mechanisms. Future research will focus on optimizing bounding box localization, integrating multi-source remote sensing data, and engineering the deployment of the algorithm on embedded edge devices, with the aim of further enhancing the algorithm's overall performance in practical operational scenarios.

References

[1] D. Lv, C. Zhao, H. Ye, et al.: GS-YOLO: A Lightweight SAR Ship Detection Model Based on Enhanced GhostNetV2 and SE Attention Mechanism, IEEE Access, Vol.12 (2024), p. 108414-108424.

[2] D. Pan, Y. Wu, W. Dai, et al.: TAG-Net: Target Attitude Angle-guided Network for Ship Detection and Classification in SAR Images, Remote Sensing, Vol. 16 (2024) No.6, p.944.

[3] D. Pan, X. Gao, W. Dai, et al.: SRT-Net: Scattering Region Topology Network for Oriented Ship Detection in Large-scale SAR Images, IEEE Transactions on Geoscience and Remote Sensing, Vol. 62 (2024), p.1-18.

[4] M. Keymasi, O. Ghozatlou and M. Datcu: Goal-oriented Semantic Modules for SAR Ship Detection, Proc. International Workshop on the Theory of Computational Sensing and Its Applications to Radar, Multimodal Sensing and Imaging (CoSeRa) (2024), p.16-20.

[5] S. Zhou, M. Zhang, L. Wu, et al.: Lightweight SAR Ship Detection Network Based on Transformer and Feature Enhancement, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 17 (2024), p.4845-4858.

[6] B. Pongrac and D. Gleich: Despeckling of SAR Images Using Residual Twin CNN and Multi-resolution Attention Mechanism, Remote Sensing, Vol. 15 (2023) No.14, p.3698.

[7] S. Liu, S. Tian, Y. Zhao, et al.: LG-DBNet: Local and Global Dual-branch Network for SAR Image Denoising, IEEE Transactions on Geoscience and Remote Sensing, Vol. 62 (2024), p.1-15.

[8] S. Tian, G. Jin, J. Gao, et al.: Ship Detection in Synthetic Aperture Radar Images Based on Bilevel Spatial Attention and Deep Poly Kernel Network, Journal of Marine Science and Engineering, Vol. 12 (2024) No.8, p.1379.

[9] J. Li, Z. Yu, J. Chen, et al.: Improve the Performance of SAR Ship Detectors by Small Object Detection Strategies, Remote Sensing, Vol. 16 (2024) No.17, p.3338.

[10] L. Ying, D. Miao and Z. Zhang: A Robust One-stage Detector for SAR Ship Detection with Sequential Three-way Decisions and Multi-granularity, Information Sciences, Vol. 667 (2024), p.120436.

[11] C.Q. Zhang, Y. Deng, M.Z. Chong, et al.: Entropy-Based Re-sampling Method on SAR Class Imbalance Target Detection, ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 209 (2024), p.432-447.

[12] Z. He, K. Wang, T. Fang, et al.: Comprehensive Performance Evaluation of YOLOv11, YOLOv10, YOLOv9, YOLOv8 and YOLOv5 on Object Detection of Power Equipment, Proc. Chinese Control and Decision Conference (CCDC) (2025), p.1281-1286.

[13] X. Ren, Y. Bai, G. Liu, et al.: YOLO-lite: An Efficient Lightweight Network for SAR Ship Detection, Remote Sensing, Vol. 15 (2023) No.15, p.3771.