ICJK
Chinese Journal of Information Fusion
http://dx.doi.org/10.62762/CJIF.2025.869982

RESEARCH ARTICLE

# An Improved Yolov12-Based Object Detection Model For Ship Monitoring in SAR Images

**Wenqi Wang**[1], **Xu Zhang**[1,*], **Jun Ma**[1], **Xunhuan Ren**[1] **and Viktar Yurevich Tsviatkou**[1]

[1] Department of Infocommunication Technologies, Belarusian State University of Informatics and Radioelectronics, Minsk 220013, Belarus

## Abstract

**Ship detection in Synthetic Aperture Radar (SAR) imagery is crucial for maritime surveillance. However, it faces significant challenges, including small target sizes, complex sea clutter interference, and stringent requirements for computational efficiency in on-board processing. While detection frameworks like YOLOv12 have achieved a favorable balance between speed and accuracy by integrating attention mechanisms with convolutional neural networks (CNNs), their generic architectures are not optimized for the unique physical characteristics of SAR imagery and the scattering properties of ship targets. To develop a more suitable lightweight and high-precision model for SAR ship detection, this study proposes an improved YOLOv12 framework. Specifically, two modules are adopted: First, the GhostStem module is embedded into the shallow network layers to replace traditional convolutional layers. This lightweight feature extraction module effectively reduces the number of parameters and computational cost in the early stages, establishing an efficient foundation for target detection in SAR images. Second, the OverLookGate (OLGate) module is incorporated. By extracting lightweight global semantic priors and employing a two-level feature gating mechanism, it significantly enhances the model's capability to discriminate and localize features within SAR imagery under complex backgrounds (e.g., coastlines and island interference) and among distributed small-scale ship targets. Experiments on publicly available SAR ship detection datasets show that, compared with the original YOLOv12 and other mainstream detectors, the proposed improved model maintains high accuracy while demonstrating competitive performance, particularly achieving significant improvements in Recall and mAP@0.5, especially in achieving higher recall and overall accuracy for small targets in complex scenarios.**

**Keywords**: YOLOv12, ship detection, small-scale target, synthetic aperture radar (SAR).

## 1 Introduction

Synthetic Aperture Radar (SAR), with its stable all-weather, all-day observation capability, has

**Academic Editor:**
Dongdong Li

**Submitted:** 29 December 2025
**Accepted:** 04 June 2026
**Published:** 11 June 2026

**Vol.** 3, **No.** 2, 2026.
10.62762/CJIF.2025.869982

***Corresponding author:**
✉ Xu Zhang
xuz49362@gmail.com

**Citation**
Wang, W., Zhang, X., Ma, J., Ren, X., & Tsviatkou, V. Y. (2026). An Improved Yolov12-Based Object Detection Model For Ship Monitoring in SAR Images. *Chinese Journal of Information Fusion*, 3(2), 125–137.

© 2026 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

become an indispensable data source for maritime vessel detection and surveillance [1]. Compared to optical remote sensing, SAR is insensitive to clouds, rain, fog, and lighting conditions, enabling reliable acquisition of sea surface information [2]. In SAR imagery, metal hulls appear as bright spots due to their strong backscattering characteristics, providing a physical basis for ship detection. This technology has been implemented in operational systems (e.g., the EU's CleanSeaNet [3]) and works synergistically with the Automatic Identification System (AIS) [4], effectively supporting maritime safety tasks such as traffic management and monitoring illegal activities like locating vessels with AIS switched off. It is thus a crucial component within current maritime monitoring frameworks, which increasingly integrate spaceborne optical and radar sensors for comprehensive vessel surveillance [5].

However, SAR-based ship detection still faces a series of technical challenges: the imagery is susceptible to interference from speckle noise and sea surface wave clutter; strong land echoes in nearshore areas lead to a high false alarm rate; and small vessel targets have a low pixel ratio and weak features, making them difficult to detect. Precisely because of these inherent difficulties, while SAR ship detection technology is relatively mature, there remains significant room for performance improvement. To address these challenges and enhance the robustness and generalization capability of detection models across different scenarios (e.g., complex sea states, ports, and coastal waters), current research trends frequently involve leveraging high-resolution optical imagery for algorithm validation and performance enhancement, and employing multi-source data fusion to compensate for the limitations of SAR data alone.

At present, SAR image target detection methods can be mainly divided into two categories: traditional methods and deep learning methods, with anchor-free deep learning approaches representing a recent development in the latter category [6]. Traditional methods are usually based on the physical scattering characteristics or image statistical features of ship targets, such as using constant false alarm rate (CFAR) detectors and their variants [7], which have good detection efficiency in uniform sea clutter background. The latter, including methods like AdaBoost [8] and Support Vector Machines (SVM) [9], often serve as post-processing classifiers. These classifiers utilize manually designed features (e.g., geometric, texture, or scattering features) to distinguish true targets from false alarms or to perform basic categorization. However, these methods are limited in complex scenes: the images are easily affected by speckle noise and sea surface clutter; due to strong land echoes, the false alarm rate in nearshore areas is very high; traditional methods are difficult to detect small ship targets with low pixel ratio and weak features; in addition, SAR images lack the fine texture and structural information of ships, resulting in insufficient ability of traditional methods in ship type classification.

The rapid development of deep learning, particularly the success of Convolutional Neural Networks (CNNs), has revolutionized object detection in remote sensing imagery, enabling the generation of task-specific anchors and feature representations for small ship targets [10]. The powerful feature learning capability of CNNs has replaced traditional manual feature design, with patch-based deep feature representations significantly improving classification accuracy and efficiency in high-resolution SAR imagery [11]. Currently, the YOLO series of models, due to their efficient detection framework, have been widely applied to SAR image interpretation, such as YOLOv4 [12] and YOLOv8 [13]. However, YOLO architectures, which continuously strive for a better speed-accuracy balance, still demonstrate insufficient feature discrimination and generalization capabilities when faced with the unique challenges of SAR ship detection–such as strong speckle noise, complex sea clutter, and large variations in target scale. Simultaneously, the design of general-purpose models often struggles to meet the stringent lightweight requirements of edge platforms like satellite-borne or shipborne systems.

To develop a SAR ship detection model that is both lightweight and high-precision, this study selects the latest YOLOv12 [14] as the baseline framework. This choice is based on two primary considerations: first, YOLOv12 represents the current performance benchmark for real-time object detection, and its "attention-centric" design achieves an outstanding speed-accuracy balance on general datasets; second, its architecture is highly modular with good extensibility, facilitating the integration of targeted optimization mechanisms. However, as a general-purpose detector, YOLOv12 is not designed for the unique physical characteristics of SAR imagery, such as strong speckle noise, dense small target distributions, and complex nearshore backgrounds. Its standard convolutional layers are inefficient at suppressing SAR-specific noise, and their feature discriminability for densely packed

small targets is limited, leading to a clear performance bottleneck when applied directly.

Therefore, we propose a lightweight and high-precision detection framework based on the improved YOLOv12. The core innovation lies in the introduction of two collaborative modules: the GhostStem lightweight feature extractor and the OLGate global-local feature gating module. By integrating these two modules into the YOLOv12 model, the efficiency of SAR ship detection in complex backgrounds is enhanced while maintaining high accuracy. The main contributions of this paper are summarized as follows:

- The GhostStem lightweight module was designed to replace the traditional shallow convolutional layers in the network. This effectively reduced the computational load in the early stage of the model, laying a solid foundation for efficient SAR image target detection.
- The OverLookGate (OLGate) module was adopted. By extracting lightweight global semantic priors and employing a two-level feature gating mechanism, the model's ability to extract SAR image features in complex backgrounds and small target scenarios was significantly enhanced.
- Comprehensive experiments validate the effectiveness of our approach. Compared to the original YOLOv12 and other mainstream detectors, the proposed improved model maintains high accuracy while significantly improving the mean average precision (mAP), particularly excelling in complex scenarios and small target detection. This demonstrates its practical value for real-world applications.

## 2 Related Work

Ship detection in Synthetic Aperture Radar (SAR) imagery plays a pivotal role in maritime surveillance. In recent years, with the advancement of deep learning, researchers have made significant efforts to balance detection accuracy with inference efficiency. Existing SAR ship detection methods can be broadly categorized into two streams: one focuses on high-performance detection tailored for small targets and complex backgrounds, which often yields high accuracy but incurs substantial computational costs; the other prioritizes lightweight design for edge deployment, which achieves fast inference but frequently suffers from accuracy degradation due to insufficient feature extraction capabilities when dealing with the weak textures and strong noise inherent in SAR images.

### 2.1 High-Performance Detection Methods for Small Targets and Complex Scenarios

To address challenges such as large-scale variations, complex nearshore backgrounds, and weak features of small targets in SAR imagery, researchers have proposed various complex detection frameworks and feature enhancement strategies.

Regarding detection frameworks, anchor-based methods like Faster R-CNN [15] utilize preset multi-scale anchor boxes to accommodate target variations. However, ship scales in SAR images vary drastically–from small fishing boats to large cargo vessels. Fixed anchor designs often fail to match actual targets, leading to low recall rates for small targets and anchor conflicts in dense clusters. Conversely, anchor-free methods such as CenterNet [16] circumvent anchor design but rely heavily on precise keypoint regression. Under the interference of strong speckle noise in SAR images, keypoint localization becomes unstable, resulting in increased false alarms.

In terms of feature enhancement, strategies like BiFPN, introduced as part of the EfficientDet framework [17], add dense bidirectional cross-scale connections to capture weak small-target features. While this strengthens feature interaction, the complex multi-path structure introduces excessive parameters, and the dense connections tend to amplify sea clutter interference. Similarly, ASFF [18] attempts to fuse multi-scale features via learnable spatial weights. However, due to the significant domain gap between SAR and optical images, its weight learning process exhibits poor adaptability in complex sea conditions (e.g., high sea states, nearshore areas), failing to effectively suppress false alarms. In summary, while these methods improve accuracy, they compromise real-time performance, and some complex mechanisms prove ill-suited for the SAR domain.

### 2.2 Lightweight Detection Models and Feature Extraction Design

To meet the processing demands of spaceborne or airborne platforms, lightweight design has emerged as another critical research direction. These methods primarily reduce computational load by simplifying

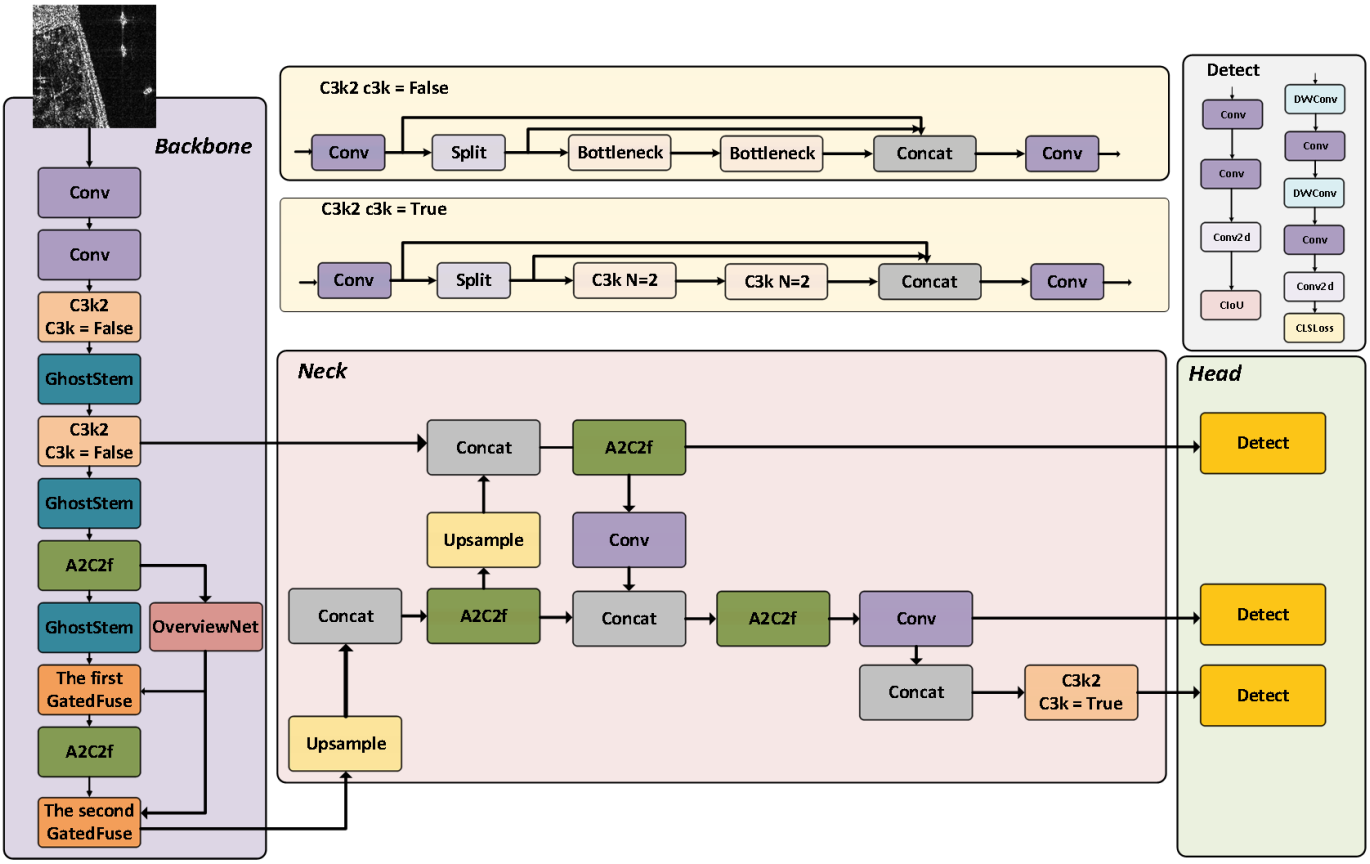



**Figure 1.** Improved YOLOv12 framework diagram.

network structures or employing lightweight operators.

Lightweight networks represented by ShuffleNet [19] promote information flow through channel shuffling. However, their heavy reliance on grouped convolutions and frequent channel rearrangement disrupts the local correlations within feature map. This is particularly detrimental to SAR ship detection, as the identification of small-scale ships depends heavily on coherent local pixel relationships, and such disruption often leads to missed detections. Regarding attention mechanisms, while the SE module [20] effectively recalibrates channel features with minimal overhead, it neglects the spatial dimension. Given that SAR ships often follow specific spatial distribution priors (e.g., linear arrangements along shipping lanes), relying solely on channel attention fails to leverage this spatial context to suppress scattered background clutter.

Furthermore, even YOLO-based detectors adapted for ship detection [21], while achieving improved speed, inherit architectural designs primarily optimized for natural images and do not fully address the unique backscattering characteristics of SAR imagery. Although the recently released YOLOv12 incorporates attention mechanisms, as a generic detector, it is not optimized for the strong backscattering characteristics of SAR imagery. When directly applied to SAR scenes, it struggles to maintain discrimination capability for weak, small, and dense targets while significantly compressing parameters.

In summary, although existing research has made significant progress in SAR ship detection, a core contradiction remains: high-precision models often have complex structures and high computational costs, making it difficult to meet the processing requirements of spaceborne or airborne platforms. Conversely, lightweight models, after drastic parameter reduction, struggle to effectively cope with the challenges of weak small-target features and strong background clutter in SAR images. There is currently a lack of a lightweight solution that can deeply integrate SAR domain-specific prior knowledge. To address this issue, this study proposes an improved model based on the YOLOv12 framework, incorporating the GhostStem and OLGate modules. The GhostStem module is designed to build a lightweight feature extraction front-end, reducing the computational burden from the very beginning. The OLGate module,

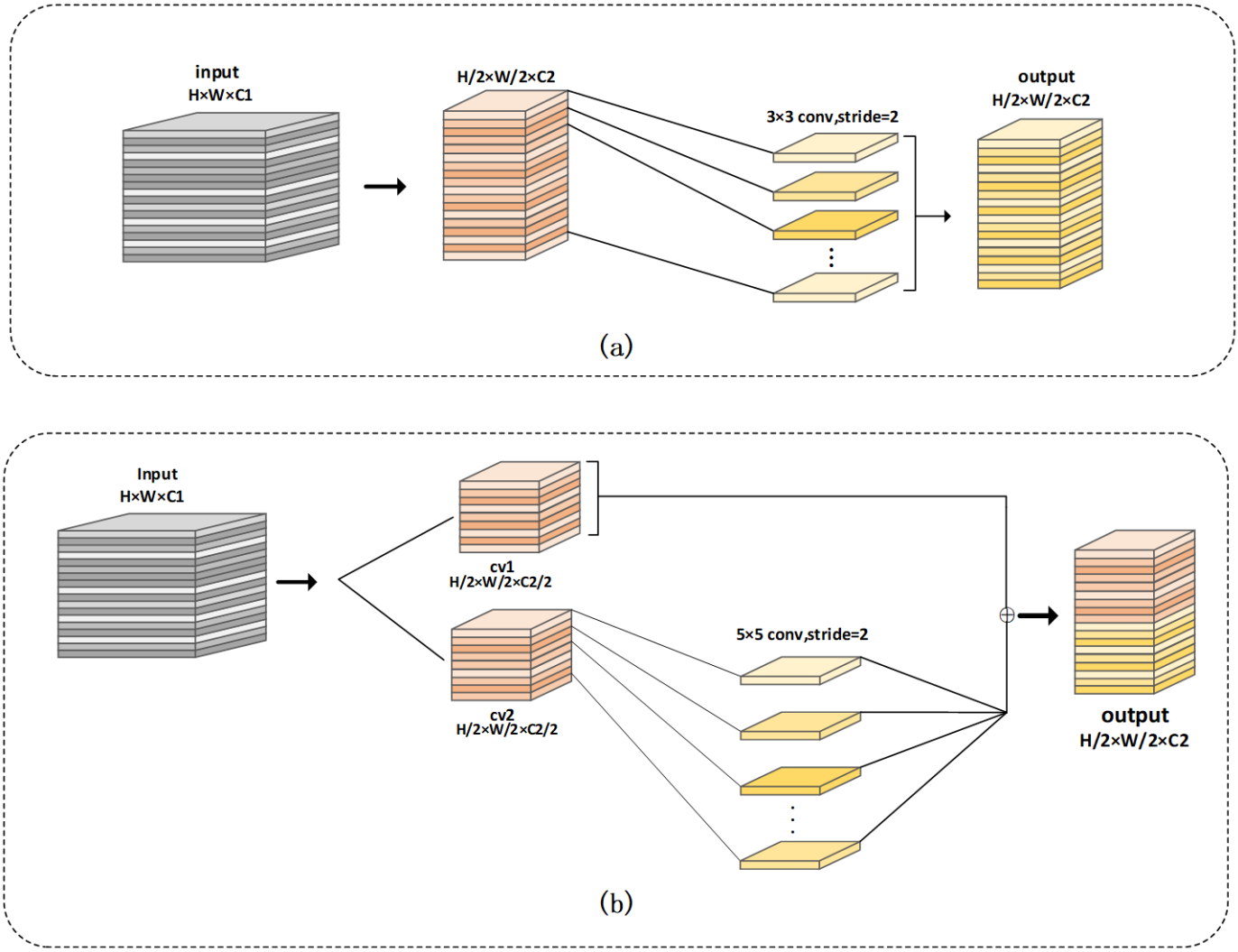



**Figure 2.** Channel number handling of Conv (a) and Ghostconv (b).

by simulating a "glance-then-examine" mechanism, enhances the model's ability to discriminate complex backgrounds and small targets. The synergistic operation of these two modules aims to fill the gaps in existing methods and achieve an optimal balance between speed and accuracy in SAR ship detection.

## 3 Methodology

The proposed improved network is based on the YOLOv12 baseline and introduces two core architectural enhancements tailored to SAR images, both deployed in the backbone (as shown in Figure 1). First, the GhostStem module is embedded in the shallow to middle layers of the backbone, replacing multiple standard convolutional layers to build a lightweight feature extraction front-end and reduce computational cost. Second, the OverLookGate (OLGate) module, consisting of OverviewNet and GatedFuse, is integrated into the middle-to-later stages of the backbone. OverviewNet extracts lightweight global semantic priors from intermediate features, while GatedFuse adaptively fuses them with subsequent deep local features. By combining a lightweight front-end with a global-local feature modulation mechanism, our network maintains high inference speed while significantly improving detection robustness for small targets and complex backgrounds in SAR images.

### 3.1 GhostStem

GhostNet, as a highly efficient and lightweight convolutional neural network architecture, aims to significantly reduce computational costs by leveraging redundancy in feature maps and employing inexpensive linear operations [22]. To better adapt to the characteristics of SAR image target detection tasks, which often involve salient small targets and strong background noise, this study further improves its core mechanism and encapsulates it into the GhostStem module, serving as a plug-and-play

lightweight network entry point (Stem layer). The core computation of this module is performed by the internally integrated GhostConv submodule, which inherits the core idea of utilizing redundancy and inexpensive operations. However, by adjusting the structure (using dual convolutions and a fixed $5 \times 5$ kernel) and positioning (as a dedicated Stem), it achieves both lightweighting and enhanced early capture capabilities of key information in SAR images. This enables efficient feature extraction and preliminary enhancement at the network front end, thereby improving the model's overall ability to capture SAR image features and its inference efficiency.

The design of GhostConv is based on the observation of feature map redundancy (as shown in Figure 2). Its core idea is to decompose the standard single convolution into two stages: first, a set of intrinsic feature maps is generated through a small number of convolutions; then, these features are expanded into richer phantom feature maps, thereby significantly reducing the number of parameters and computational complexity while maintaining the diversity of feature representation. Specifically, the structure of GhostConv is as follows: For a feature map with $c_1$ input channels and $c_2$ target output channels, a convolutional layer $cv1$ is first used to compress the number of channels to $c_2/2$, completing the extraction of basic features and preliminary dimensionality reduction. Subsequently, the obtained intermediate features are input to a second convolutional layer $cv2$. In this study, $cv2$ uses a fixed $5 \times 5$ convolutional kernel. The GhostStem module encapsulates the aforementioned GhostConv operation into a standard network layer, typically configured with a convolutional kernel $k = 3$ and a stride $s = 2$, to achieve initial downsampling and lightweight feature extraction of the input SAR image. Specifically, the outer GhostStem wrapper uses a $3 \times 3$ convolution with stride 2 for initial downsampling, while the internal cv2 submodule uses a $5 \times 5$ depthwise convolution to capture larger receptive field patterns for SAR-specific feature extraction. Placing this lightweight computational unit at the front end of the network significantly reduces the computational load during early processing of the input image. This design, which uses slightly stronger operations ($5 \times 5$ convolutions) within a lightweight framework to achieve better feature representation capabilities, is tailored to the characteristics of SAR images (small targets, strong noise).

### 3.2 OverLookGate (OLGate)

A lightweight, plug-and-play biomimetic attention module, OLGate, which employs a "glance-then-examine" mechanism is introduced. It enhances feature representations with extremely low computational overhead by extracting global semantic vectors and performing multi-level gated modulation. This design is inspired by the recent work OverLoCK, which proposes a ConvNet backbone with "glance-examine" branches and achieves spatial semantic modulation through context-mixed dynamic convolutions [23].

OLGate achieves a similar goal in a lighter way: it compresses semantic priors into global vectors rather than spatial feature maps and uses attention gating for multi-scale modulation, resulting in a more computationally efficient and easily integrated plug-in design. To adapt to the characteristics of small targets and strong noise in SAR images, OLGate has undergone targeted optimizations: it uses depthwise separable convolutions to extract noise-resistant semantic features; it deploys gated fusion at deep layers and the feature pyramid to enhance the cross-scale consistency of small targets; and it can be flexibly embedded into existing networks to support real-time SAR image processing on the edge. The OLGate module consists of an OverviewNet subnetwork and two levels of GatedFuse gating modulation units. Lightweight OverviewNet rapidly acquires image-level semantic understanding, and then dynamic gating injects this prior knowledge into feature processing flows at different scales, achieving effective fusion of global and local information (as shown in Figure 3).

OverviewNet, as the module's global semantic extractor, employs a minimalist design to achieve efficient context modeling. It receives the feature map (512 channels, $H/16 \times W/16$ resolution) output from layer C4 of the backbone network. This feature map already contains mid-level semantic information but retains a significant amount of spatial detail. The resolution is reduced to $H/32 \times W/32$ using a $3 \times 3$ depthwise separable convolution with a stride of 2, with computational cost only $1/9$ of that of a standard convolution. A $1 \times 1$ convolution is used to compress the number of channels to 128, focusing on extracting task-relevant semantic vectors. Image-level statistical features are obtained through global average pooling, and then a 128-dimensional semantic prior vector $P$ is generated through a fully connected layer.

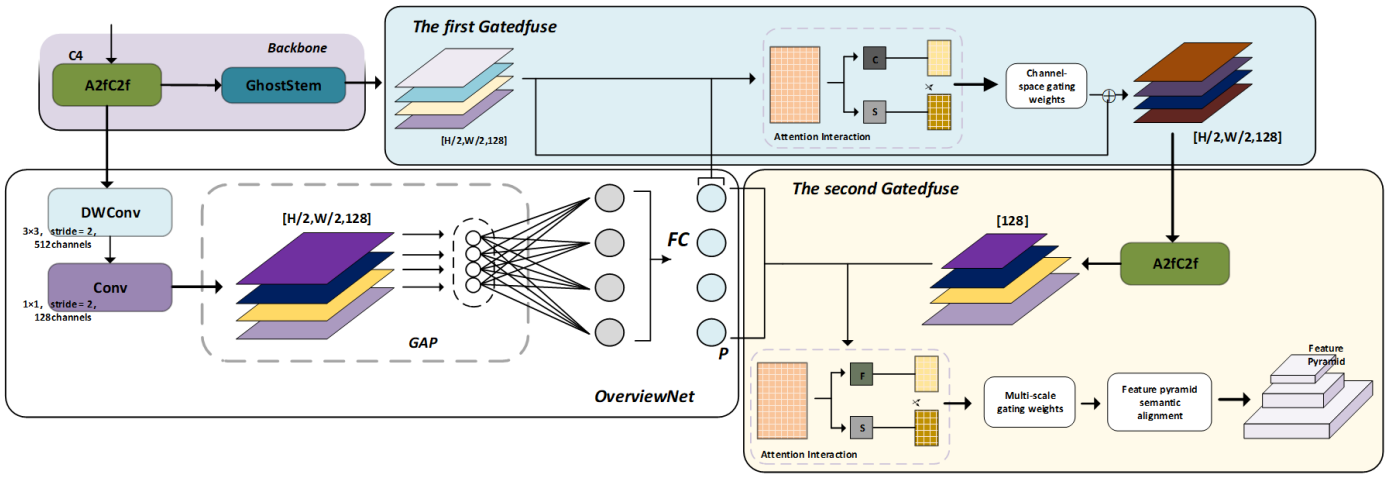



**Figure 3.** Schematic diagram of the OLGate module structure.

Mathematical representation:

$$P = \mathrm{FC}\left(\mathrm{GAP}\left(\mathrm{Conv}_{1\times1}\left(\mathrm{DWConv}_{3\times3}^{s=2}\left(F_{C4}\right)\right)\right)\right), \tag{1}$$

where $F_{C4}$ represents the output features of layer C4, DWConv represents depthwise separable convolution, GAP represents global average pooling, and FC represents a fully connected layer.

The GatedFuse module dynamically modulates feature maps using semantic priors, employing a gating mechanism to selectively enhance task-relevant information. It is strategically integrated at two critical stages: to guide deep feature extraction and to ensure semantic alignment within the feature pyramid. Operating on an attention-based mechanism, GatedFuse interacts between the global semantic prior and local feature map to generate adaptive modulation weights, which can be applied spatially or channel-wise. The process is fundamentally expressed as:

$$F_{out} = F_{in} + G(P, F_{in}) \odot T(F_{in}), \tag{2}$$

where $G$ outputs $[0, 1]$ gating values to adaptively add the transformed features $T(F_{in})$ to the input $F_{in}$, with a residual connection for stability.

The GatedFuse module operates through a two-level gating architecture to achieve precise feature modulation. The first level, Deep Feature Guidance, is integrated at the entry point of the $C5$ processing stage. It modulates the downsampled mid-level features through the following steps: semantic priors are spatially broadcast to align with the feature map dimensions; the feature map and priors are concatenated; a $1 \times 1$ convolution generates initial gating weights; a Sigmoid activation converts these weights into modulation coefficients within the $[0, 1]$ range; and a residual modulation is applied to preserve the original information flow.

The second level, Multi-scale Semantic Alignment, is deployed after the $C5$ output and prior to feature pyramid fusion. To maintain computational efficiency while handling high-channel features, this stage employs grouped convolutions. It first obtains channel-wise statistics via global average pooling, which are then fused with the semantic priors to generate channel-level gating weights. Finally, these weights perform channel-selective enhancement on the original features, ensuring semantic consistency across different scales.

## 4 Experiments

### 4.1 Dataset Description

The experiments were conducted on two public SAR ship detection datasets: RSDD-SAR and SAR-Ship-Dataset. RSDD-SAR comprises 7,000 image tiles with 10,263 ship instances, featuring multi-source, multi-mode, multi-polarization, and multi-resolution characteristics. The dataset includes 84 GF-3 tiles, 41 TerraSAR-X tiles, and 2 large uncropped images. SAR-Ship-Dataset was constructed by SAR experts using 102 Gaofen-3 and 108 Sentinel-1 images. It contains 43,819 ship chips. The hardware and software configuration is summarized in Table 1.

We implemented the proposed method alongside several popular YOLO-series models on the above platform. All models were first trained and evaluated

**Table 1.** Experimental configuration.

| Category | Attribute | Attribute Value |
|---|---|---|
| Hardware | CPU | Radeon 860M |
| | GPU | NVIDIA GeForce RTX5060 |
| | Running memory | 8.0 GB |
| | Storage memory | 256.0 GB |
| Software | IDE | PyCharm |
| | Interpreter | Python 3.12 |

**Table 2.** Hyper-parameters on SAR-Ship-Dataset.

| Attribute | Attribute Value |
|---|---|
| epochs | 300 |
| Batch size | 160 |
| imgsz | 256 |
| scale | 0.5 |
| mosaic | 1.0 |
| workers | 0 |
| optimizer | Auto |
| patience | 0 |

on the SAR-Ship-Dataset, whose training parameters are listed in Table 2. Subsequently, to assess their transferability, we fine-tuned these pre-trained models on the RSDD-SAR dataset and evaluated them again. The parameters used for this transfer learning stage are detailed in Table 3.

**Table 3.** Hyper-parameters on RSDD-SAR Dataset.

| Attribute | Attribute Value |
|---|---|
| epochs | 60 |
| Batch size | 96 |
| imgsz | 640 |
| scale | 0.5 |
| mosaic | 1.0 |
| workers | 0 |
| optimizer | Auto |
| patience | 0 |

Four standard metrics, including Precision (P), Recall (R), $mAP_{50}$, and $mAP_{50-95}$ are used to evaluate the performance. Their mathematical formulations are presented in the following.

$$P = \frac{TP}{TP + FP}, \tag{3}$$

where $TP$ is true positive, $FP$ is false positive.

$$R = \frac{TP}{TP + FN}, \tag{4}$$

where $FN$ is false negative.

$$mAP_{50} = \frac{1}{C}\sum_{c=1}^{C} AP_c\,(IoU = 0.5)\,, \tag{5}$$

where $C$ is object categories, and $AP_c$ is average precision for class $c$.

$$mAP_{50-95} = \frac{1}{10}\sum_{k=0}^{9} mAP_{0.5+0.05k}, \tag{6}$$

where $k$ is the summation index corresponding to the IoU threshold $0.5 + 0.05 \times k$, $(k = 0, 1, \ldots, 9)$.

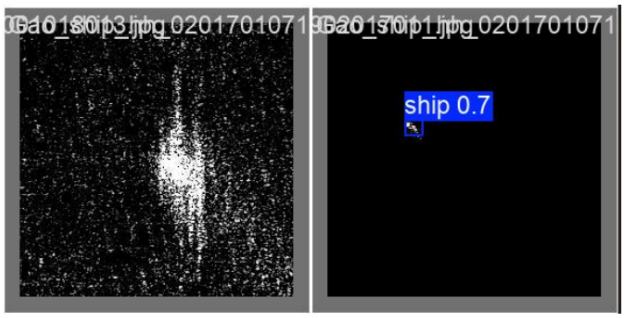
### 4.2 Visual Comparative Study

A visual comparison between the YOLOv12 model and the proposed framework was first conducted on three representative images from SAR-Ship-Dataset, whose results are presented in Figure 4. It is clearly observable that the proposed approach achieved better results than the YOLOv12 model, as it can detect more ships in noisy backgrounds. Specifically, in all three test images, the YOLOv12 model failed to detect some ships, whereas the proposed method successfully identified all potential ships without any omissions. On the other hand, the YOLOv12 model sometimes produced redundant bounding boxes, such as misidentifying one ship as two adjacent ships, as shown in Figure 4(c), while the proposed method still generated high-quality bounding boxes with accurate localization.

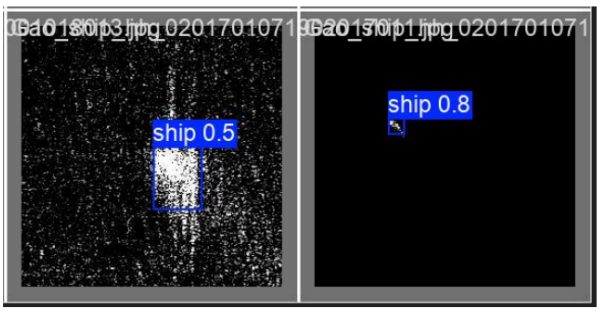
This result demonstrates that the proposed method significantly enhances the understanding of the overall structure of ship targets through improved feature fusion and target perception mechanisms. Thereby, it avoids common issues in YOLOv12, such as overlapping bounding boxes and incorrect segmentation, and achieves more reliable and complete target detection.

### 4.3 Comprehensive Comparative Study on SAR-Ship-Dataset

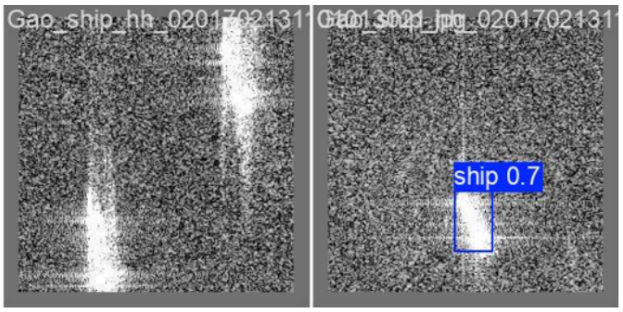
To evaluate the performance of the proposed method comprehensively, we conducted experiments on the full SAR-Ship-Dataset using its default train/validation/test split. For comparison, several representative YOLO-series models were also trained and tested under the same settings, including YOLOv5 [24], YOLOv6 [25], YOLOv8 [13], LEAD-YOLO [26], and YOLOv12 [14]. All models were evaluated on the independent test set using



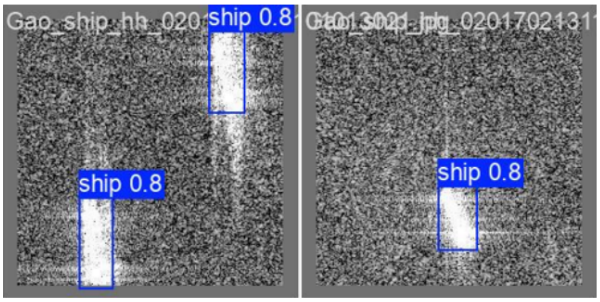

Ship Detection using YOLOv12 on Img (a)



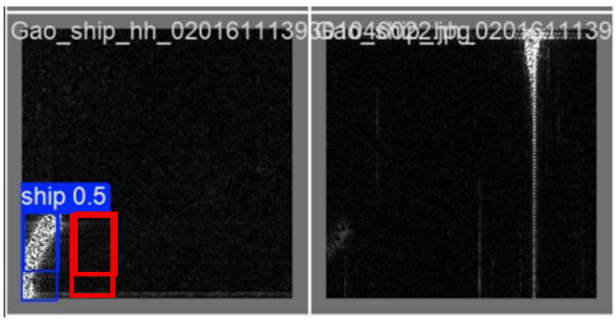

Ship Detection using Proposed method on Img (a)



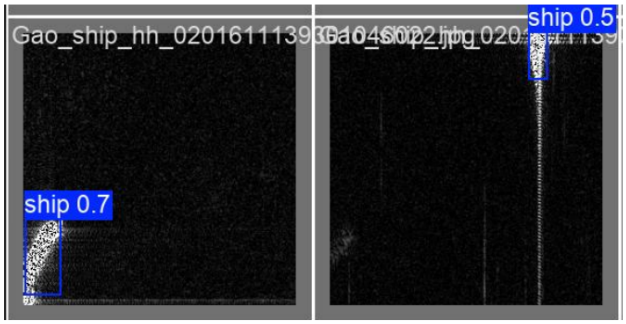

Ship Detection using YOLOv12 on Img (b)




Ship Detection using Proposed method on Img (b)




Ship Detection using YOLOv12 on Img (c)




Ship Detection using Proposed method on Img (c)

**Figure 4.** Visual comparison of YOLOv12 (left) and the proposed method (right) on three SAR images from the SAR-Ship-Dataset.

metrics mentioned in Sec. 4.1. The comparative results are summarized in Table 4.

**Table 4.** Comparative Results on the test set of the SAR-Ship-Dataset.

| METHODS | PRECISION (%) | Recall (%) | $mAP_{50}$ | $mAP_{50-95}$ |
|---|---|---|---|---|
| YOLOv5 | 88.5 | 85.5 | 94.0 | 54.8 |
| YOLOv6 | 89.9 | 87.4 | 93.6 | 55.8 |
| YOLOv8 | 92.3 | 88.5 | 95.7 | 55.0 |
| LEAD-YOLO | 93.5 | 86.1 | 93.4 | 57.2 |
| YOLOv12 | 87.4 | 83.8 | 92.4 | 51.2 |
| Proposed method | 92.7 | 90.5 | 96.1 | 54.5 |

From Table 4, it can be seen that the proposed method achieves the highest Recall (90.5%) and mAP50 (96.1%) among all evaluated methods, indicating strong completeness in detecting targets and competitive overall performance under a standard intersection over union (IoU) threshold. While its Precision (92.7%) is slightly lower than LEAD-YOLO's 93.5%, it remains very competitive. However, its performance on the stricter localization metric mAP50-95 is 54.5%, trailing behind LEAD-YOLO (57.2%). Overall, the proposed method demonstrates a notable improvement in detection recall and overall accuracy, while its localization precision under stricter IoU thresholds remains competitive and offers room for further refinement.

### 4.4 Comparative Study on RSDD-SAR Dataset

For further investigating the generalization of the proposed method, we conduct transfer learning on RSDD-SAR dataset and evaluate the implemented models from the perspective of precision, recall, mAP50 and mAP50-95. The comparative result is shown in Table 5.

**Table 5.** Comparisons with the baseline model using RSDD-SAR.

| METHODS | PRECISION (%) | Recall (%) | mAP50 (%) | mAP50-95 (%) |
|---|---|---|---|---|
| YOLOv5 | 92.3 | 88.5 | 95.7 | 55.8 |
| YOLOv6 | 90.5 | 90.0 | 95.5 | 56.0 |
| YOLOv8 | 91.6 | 88.9 | 95.6 | 57.5 |
| LEAD-YOLO | **94.1** | 86.1 | 93.9 | 59.1 |
| YOLOv12 | 93.1 | 87.0 | 94.4 | **67.9** |
| Proposed method | 93.7 | **90.8** | **96.9** | 58.2 |

From Table 5, it is clear that the proposed method delivers a strong and balanced overall performance. It achieves the highest scores in both Recall, which is 90.8%, and mAP50, which is 96.9%, indicating its superior capability in identifying nearly all target ships and its top-ranked general detection accuracy under the standard IoU threshold. While its Precision of 93.7% is excellent and very close to the best-performing model (LEAD-YOLO at 94.1%), the proposed method does not lead in the most stringent localization metric, mAP50-95 (58.2%). This metric is notably dominated by YOLOv12 (67.9%), suggesting that YOLOv12 excels at generating bounding boxes with extremely precise coordinates, whereas the proposed method and others like LEAD-YOLO show competitive but more moderate performance in this aspect. In summary, the proposed method offers the best trade-off for reliable and complete detection, excelling in finding targets and overall accuracy, with room for potential refinement in precise boundary box regression compared to the specialized strength of YOLOv12.

### 4.5 Analysis of Model Complexity and Efficiency

To analyze the model complexity of the proposed model, the number of parameter and floating-point operations (FLOPs) of our model is compared with different type of YOLOv12 models. On the other hand, we use metric of the efficiency-improvement ratio, which is defined as the ratio of relative mAP gain to relative parameter increase, to quantify the design efficiency of the model.

**Table 6.** Lightweight performance comparison: Proposed method vs. YOLOv12 variants.

| METHODS | params (M) | FLOPs (G) | Efficiency improvement ratio |
|---|---|---|---|
| YOLOv12n | 2.6 | 6.0 | 1.00 (baseline) |
| YOLOv12s | 9.3 | 19.4 | 0.34 |
| YOLOv12m | 20.2 | 59.8 | 0.18 |
| YOLOv12l | 26.4 | 82.4 | 0.14 |
| YOLOv12x | 59.1 | 184.6 | 0.09 |
| Proposed method | 3.4 | 7.9 | 1.44 |

Table 6 reveals the superior performance-to-complexity trade-off achieved by the proposed model. It attains the highest Efficiency Improvement Ratio of 1.44, surpassing the YOLOv12n baseline value of 1.00 by a significant 44 percent. This is accomplished with only a modest increase in model size, as the proposed model requires 3.4 million parameters and 7.9 GFLOPs compared to the baseline's 2.6 million parameters and 6.0 GFLOPs. The table clearly demonstrates the steep trade-off efficiency within the YOLOv12 family itself, where larger variants like YOLOv12x exhibit massive growth in parameters and FLOPs but suffer a drastic reduction in their efficiency ratios, with YOLOv12x falling to 0.09. Therefore, our model successfully positions itself as a highly efficient architecture, delivering a more favorable balance between computational cost and effective capacity than the scaled variants of the contemporary YOLOv12 benchmark.

### 4.6 Ablation Experiments

To verify the effectiveness of each component in proposed model, we conducted ablation experiments on the SAR-Ship-Dataset by successively removing the OLGate module and the GhostStem module. The results are shown in Table 7.

Table 7 clearly demonstrates the individual and combined contributions of the proposed modules. When only the GhostStem module is enabled, the model achieves moderate precision but suffers from the lowest recall, indicating a tendency to miss targets. Conversely, enabling only the OLGate module yields a significant boost in recall, the highest among all ablative configurations, at the cost of a slight dip in precision. This suggests the OLGate module is crucial for comprehensive feature aggregation and reducing missed detections. The complete proposed model, integrating both the GhostStem and OLGate modules, achieves the best performance across all key metrics: precision, recall, mAP50, and mAP50-95. This synergistic improvement confirms that GhostStem effectively enhances feature representation and localization precision, and when combined with OLGate's superior recall capability, results in a more robust and accurate detector. The progressive increase in mAP50-95 further validates that the full model provides better localization accuracy under stricter criteria.

**Table 7.** Comparison of ablation test results.

| GHOSTSTEM | OLGate | Precision (%) | Recall (%) | mAP50 (%) | mAP50-95 (%) |
|---|---|---|---|---|---|
| Yes | No | 89.7 | 84.8 | 93.5 | 50.5 |
| No | Yes | 88.9 | 90.2 | 94.5 | 52.7 |
| Yes | Yes | 92.7 | 90.5 | 96.1 | 54.5 |

## 5 Conclusion

This study addresses the persistent challenges in SAR ship detection—small targets, complex sea clutter, and the demand for efficient on-board processing—by proposing a lightweight and high-precision model based on an improved YOLOv12 framework. In contrast to general-purpose detectors and existing specialized approaches reviewed in the related work, our solution introduces two dedicated modules: the GhostStem for efficient shallow feature extraction and the OverLookGate (OLGate) for extracting lightweight global semantic priors and employing a two-level feature gating mechanism. This design explicitly targets the unique scattering characteristics of SAR ships and complex background interference, areas where prior methods exhibit limitations in efficiency, adaptability, or feature representation.

Experimental results validate the effectiveness of our approach. On the SAR-Ship-Dataset, our model achieves competitive performance in core metrics. It attains the highest Recall (90.5%) and mAP50 (96.1%), demonstrating strong capability in identifying ship targets and overall detection accuracy within complex SAR scenes. Simultaneously, it maintains high Precision (92.7%), indicating robust control over false alarms. These outcomes confirm that the proposed GhostStem and OLGate modules successfully enhance the model's feature discrimination and localization capabilities for small targets, as intended.

However, the experiments also reveal specific areas for future improvement. The model's performance on the more stringent localization metric mAP50-95 under high IoU thresholds can be further enhanced. This points towards potential optimization in the model's boundary refinement process.

In summary, this research presents a balanced and effective solution for SAR ship detection. By integrating task-specific lightweight designs, the proposed model notably improves detection accuracy, particularly for small targets, while maintaining the efficiency crucial for applications. Future work will focus on refining the localization mechanism to achieve more precise bounding box predictions.

## Data Availability Statement

The datasets used in this study are publicly available. The SAR-Ship-Dataset is available at https://github.com/CAESAR-Radi/SAR-Ship-Dataset. The RSDD-SAR dataset is available at https://github.com/makabakasu/RSDD-SAR-OPEN.

## Funding

This work was supported without any funding.

## Conflicts of Interest

The authors declare no conflicts of interest.

## AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

## Ethical Approval and Consent to Participate

Not applicable.

## References

[1] Maître, H. (Ed.). (2013). *Processing of synthetic aperture radar (SAR) images*. John Wiley & Sons.

[2] Chan, Y. K., & Koo, V. (2008). An introduction to synthetic aperture radar (SAR). *Progress In Electromagnetics Research B*, *2*, 27-60. [CrossRef]

[3] Carpenter, A. (2015). European maritime safety agency CleanSeaNet activities in the North Sea. In *Oil Pollution in the North Sea* (pp. 33-47). Cham: Springer International Publishing. [CrossRef]

[4] Svanberg, M., Santén, V., Hörteborn, A., Holm, H., & Finnsgård, C. (2019). AIS in maritime research. *Marine Policy*, *106*, 103520. [CrossRef]

[5] Kanjir, U., Greidanus, H., & Oštir, K. (2018). Vessel detection and classification from spaceborne optical images: A literature survey. *Remote sensing of environment*, *207*, 1-26. [CrossRef]

[6] Sun, Z., Dai, M., Leng, X., Lei, Y., Xiong, B., Ji, K., & Kuang, G. (2021). An anchor-free detection method

for ship targets in high-resolution SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14*, 7799-7816. [CrossRef]

[7] Li, T., Liu, Z., Xie, R., & Ran, L. (2017). An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11*(1), 184-194. [CrossRef]

[8] Wei, G., Qingwen, Q., Lili, J., & Ping, Z. (2008, July). A new method of SAR image target recognition based on AdaBoost algorithm. In *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium* (Vol. 3, pp. III-1194). IEEE. [CrossRef]

[9] Anagnostopoulos, G. C. (2009). SVM-based target recognition from synthetic aperture radar images using target region outline descriptors. *Nonlinear Analysis: Theory, Methods & Applications, 71*(12), e2934-e2939. [CrossRef]

[10] Yue, T., Zhang, Y., Liu, P., Xu, Y., & Yu, C. (2022). A generating-anchor network for small ship detection in SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15*, 7665-7676. [CrossRef]

[11] Ren, Z., Hou, B., Wen, Z., & Jiao, L. (2018). Patch-sorted deep feature learning for high resolution SAR image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11*(9), 3113-3126. [CrossRef]

[12] Gao, Y., Wu, Z., Ren, M., & Wu, C. (2022). Improved YOLOv4 based on attention mechanism for ship detection in SAR images. *IEEE Access, 10*, 23785-23797. [CrossRef]

[13] Yasir, M., Liu, S., Pirasteh, S., Xu, M., Sheng, H., Wan, J., ... & Li, J. (2024). YOLOShipTracker: Tracking ships in SAR images using lightweight YOLOv8. *International Journal of Applied Earth Observation and Geoinformation, 134*, 104137. [CrossRef]

[14] Tian, Y., Ye, Q., & Doermann, D. (2026). Yolov12: Attention-centric real-time object detectors. *Advances in neural information processing systems, 38*, 78433-78457.

[15] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems, 28*.

[16] Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., & Tian, Q. (2019, October). CenterNet: Keypoint Triplets for Object Detection. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 6568-6577). IEEE. [CrossRef]

[17] Tan, M., Pang, R., & Le, Q. V. (2020, June). EfficientDet: Scalable and Efficient Object Detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 10778-10787). IEEE. [CrossRef]

[18] Liu, S., Huang, D., & Wang, Y. (2019). Learning spatial fusion for single-shot object detection. *arXiv preprint arXiv:1911.09516*. [CrossRef]

[19] Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018, June). ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6848-6856). IEEE. [CrossRef]

[20] Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2019). Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(8), 2011-2023. [CrossRef]

[21] Zhang, Y., Hao, L. Y., & Li, Y. (2024, December). SD-YOLO: An attention mechanism guided YOLO network for ship detection. In *2024 14th International Conference on Information Science and Technology (ICIST)* (pp. 769-776). IEEE. [CrossRef]

[22] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020, June). GhostNet: More Features From Cheap Operations. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1577-1586). IEEE. [CrossRef]

[23] Lou, M., & Yu, Y. (2025, June). OverLoCK: An Overview-first-Look-Closely-next ConvNet with Context-Mixing Dynamic Kernels. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 128-138). IEEE. [CrossRef]

[24] Jocher, G., Stoken, A., Borovec, J., NanoCode012, ChristopherSTAN, Liu, C., Laughing, Hogan, A., lorenzomammana, tkianai, yxNONG, AlexWang1900, Diaconu, L., Marc, wanghaoyang0106, ml5ah, Doug, Hatovix, Poznanski, J., Yu, L., changyu98, Rai, P., Ferriday, R., Sullivan, T., Wang, X., YuriRibeiro, Claramunt, E. R., hopesala, dave, p., & yzchen. (2020). *ultralytics/yolov5: v3.0* (v3.0) [Computer software]. Zenodo. [CrossRef]

[25] Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., ... & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*. [CrossRef]

[26] Mo, H., Wu, J., Xia, H., Yu, X., & Zhao, A. E. (2025). A lightweight, efficient, adaptive design of YOLOv5 for enhanced SAR ship detection. *Remote Sensing Letters, 16*(5), 549-559. [CrossRef]

**Wenqi Wang** enrolled in the Communication Engineering undergraduate program at Luoyang Normal University in 2022. Through the China Scholarship Council scholarship program, she began studying at the Belarusian State University of Informatics and Radioelectronics, in 2024. She is particularly interested in 5G/6G networks and AI applications in communications. (Email: wwenqi2004@gmail.com)

**Xu Zhang** enrolled in the Communication Engineering undergraduate program at Luoyang Normal University in 2022. Through the China Scholarship Council scholarship program, she began studying at the Belarusian State University of Informatics and Radioelectronics, in 2024. She is currently a senior undergraduate student. Her current research focuses on image processing and machine learning. (Email: xuz49362@gmail.com)

**Jun Ma** received the B.S. degree from the Lanzhou University of Technology, in 2015, and the M.S. degree from the Belarusian State University of Informatics and Radioelectronics, in 2018, where he is currently pursuing the Ph.D. degree. His current research interests include image processing, machine learning, and computer vision. (Email: majun@bsuir.by)

**Xunhuan Ren** received the B.S. degree from the Lanzhou University of Technology, in 2015, M.S. degree and Ph.D. degree from the Belarusian State University of Informatics and Radioelectronics, in 2018 and 2024, respectively, where she is currently an associate professor of the department of infocommunication technologies. Her current research interests include image processing and information theory. (Email: renxunhuan@bsuir.by)

**Viktar Yurevich Tsviatkou** received the Ph.D. degree from the Belarusian State University of Informatics and Radioelectronics, in 1999. He works at the Belarusian State University of Informatics and Radioelectronics, where he is currently a Professor and the Dean of the department of infocommunication technologies, Faculty of Information Security. His research interests include digital image processing, pattern recognition, signal processing, and information theory. (Email: vtsvet@bsuir.by)