DS-YOLO: A SAR Ship Detection Model for Dense Small Targets

Yifei SHEN, Qiang GAO *

College of Information Engineering, Huzhou University, 313000, Huzhou 313000, China

2023388326@stu.zjhu.edu.cn, 02794@zjhu.edu.cn

Submitted February 27, 2025 / Accepted April 29, 2025 / Online first June 11, 2025

Abstract. *Detecting dense small targets in Synthetic Aperture Radar (SAR) images has always been a challenge in ship target detection. To address this issue, this paper proposes a ship target detection model for SAR images, named DS-YOLO, which is based on the YOLO11 network architecture. The model introduces Space-to-Depth Convolution (SPDConv) module to enhance the detection capability of small targets. Additionally, a new module, Cross Stage Partial-Partial Pyramid Attention (CSP-PPA), is incorporated to improve the model's ability to extract features at multiple scales and suppress confusing backgrounds. The loss function is optimized using a bounding box loss based on Adaptive Weighted Normalized Wasserstein Distance (AWNWD), enhancing the model's adaptability to images of varying quality. Finally, experiments were conducted on the standard datasets HRSID and SAR-Ships dataset to validate the robustness and reliability of the DS-YOLO model. The experimental results show that, compared to YOLO11n, DS-YOLO achieved an mAP0.5:0.95 of 68.6% on the SAR-Ships dataset and 69.9% on the HRSID, representing improvements of 1.6% and 0.8%, respectively. Additionally, on these small-target datasets, DS-YOLO achieved an mAP0.5:0.95 of 50.8% and 60.4%, representing improvements of 4.2% and 1.2%, respectively, demonstrating higher detection accuracy.*

Keywords

DS-YOLO, SAR image, ship detection, Space-to-Depth Convolution (SPDConv), Cross Stage Partial-Partial Pyramid Attention (CSP-PPA), Adaptive Weighted Normalized Wasserstein Distance (AWNWD)

1. Introduce

Synthetic Aperture Radar (SAR), as an active microwave sensor, is less affected by illumination and weather conditions, allowing all-day, all-weather observations. Therefore, SAR-based target detection technology is widely used in both civilian and military fields, such as in the detection of ships, vehicles, aircraft, and other objects [1]. Especially in the field of maritime surveillance, using SAR satellites to detect the movements of ships on the sea surface has become a research hotspot, with its importance increasingly highlighted [2]. Traditional SAR ship detection algorithms primarily rely on the statistical distribution characteristics of sea clutter. However, these methods depend heavily on predefined distribution models, which reduces their robustness to new data and consequently limits the performance of ship detection [3]. With the rapid development of deep learning technologies and significant advancements in GPU computing power, Convolutional Neural Networks (CNN) have made revolutionary progress in the field of SAR ship detection. CNN can automatically extract useful features from complex SAR images, significantly improving detection accuracy and efficiency compared to traditional manual feature extraction methods. As a result, contemporary SAR ship detection algorithms predominantly adopt deep learning approaches [4].

Deep learning target detection algorithms, based on their anchor regression strategies, are categorized into single-stage and two-stage approaches. Single-stage detection algorithms predict the target's position and category directly on the image without a distinct region proposal stage. They are typically faster and suitable for real-time applications. Typical single-stage detection algorithms include YOLO [5], SSD [6], RetinaNet [7]. Two-stage detection algorithms, on the other hand, first generate a series of potential regions that may contain the target of interest. These potential regions are meticulously classified and their positions are further adjusted to ensure more accurate target localization. This type of algorithm generally offers higher detection accuracy but also incurs greater computational overhead. Typical two-stage detection algorithms include R-CNN [8], Fast R-CNN [9], Faster R-CNN [10] and so on.

Regarding SAR image ship target detection, many researchers globally have conducted relevant studies. For lightweight models, Lv et al. [11] proposed a regression anti-convergence loss algorithm based on a dual-regression network to address the inconsistencies between training and testing in the regression branch. This loss function allows multiple training samples in the twin regression branches to converge to the labels from opposite directions and effectively reduces parameters while improving detection accuracy through knowledge distillation. For small

DOI: 10.13164/re.2025.0407

target detection, Chai et al. [12] introduced an enhanced R-CNN algorithm by incorporating Res2Net with rich multi-scale information, constructing a bidirectional feature pyramid structure, and merging features from multiple stages of output. This approach effectively improves the detection capability for small ship targets in SAR images. To address the small-sample problem, Wang et al. [13] proposed a deep kernel learning method that maps samples to a low-dimensional embedding space using neural networks and applies kernel functions for similarity-based classification. Zhou et al. [14] introduced two novel Gaussian metric feature aggregation techniques: Gaussian Projection Distribution Metric (GPDM) and Gaussian Kernel Mean Embedding Metric (GKMEM). These techniques estimate class distributions using variational autoencoders, replacing traditional class prototypes with more robust distribution samples. They computed Wasserstein and kernel mean embedding distances to extract reliable features and proposed a Balanced Inter-class Unrelated Aggregation (BICUA) strategy. BICUA extracts support features based on sample proportions, balancing them with query features to enhance inter-class independence and reduce confusion, offering solutions to SAR data scarcity. In image processing, Zhang et al. [15] proposed a SAR ship image restoration technique based on an instance-to-image generation diffusion model. This technique repairs SAR images and generates corresponding instance-level annotations for training detection models, achieving improved accuracy. Anandhi et al. [16] proposed an enhanced approach for SAR image despeckling by integrating the Non-subsampled Contourlet Transform (NSCT) with Bayesian Maximum A Posterior (BMAP) estimation. This method effectively reduces speckle noise while preserving critical image structures, striking a balance between noise suppression and detail retention. Zhang et al. [17] proposed a SAR image-oriented ship detection network based on soft thresholding and contextual information, which effectively reduces ground noise interference. Wang et al. [18] developed a Constant False Alarm Rate (CFAR) based distance-compressed SAR data ship detector with high adaptability to sea clutter models, effectively detecting ship targets. In data augmentation, Lee et al. [19] proposed a method combining Automatic Identification System (AIS) and fishing vessel positioning system (V-Pass) information to generate training data for each ship type and developed a labeling tool to enhance the effectiveness of data generation. For the rotation box problem, Li et al. [20] proposed a new network model called TKP-Net, which utilizes rotation bounding boxes to better determine ship orientation. Benish et al. [21] proposed a matting technique for extracting targets from SAR images. Their method involves initial binary segmentation to identify target boundaries roughly, followed by trimap estimation using guided filtering. To enhance trimap accuracy, they employed super-pixel-based segmentation and applied a propagation-based matting algorithm to separate targets from the background. Experiments on MSTAR database SAR images demonstrated the effectiveness of their approach.

In object detection technologies, the YOLO series of algorithms has emerged as a research focus due to their exceptional processing efficiency and robust detection performance across various applications. Ren et al. [22] proposed an efficient lightweight network, YOLO-Lite, to improve detection efficiency for SAR ship detection. Wu et al. [23] introduced a Wavelet Cascade Residual (WCR) module based on traditional image processing techniques, specifically wavelet transform, and embedded it into an improved Spatial Pyramid Pooling (SPP) module, forming an SPP module based on wavelet transform. Feng et al. [24] proposed a lightweight position-enhanced anchor-free SAR ship detection algorithm based on YOLOX, named LPEDet. Yu et al. [25] proposed a lightweight ship detector based on YOLOX, which introduces an FPN module to achieve higher efficiency by expanding the receptive field and semantic information of single-level features, alleviating the decrease in accuracy. Wang et al. [26] proposed a new SAR ship detection method, named NAS-YOLOX, which leverages Neural Architecture Search Feature Pyramid Networks (NAS-FPN) for efficient feature fusion and multi-scale attention mechanisms for effective feature extraction. Chen et al. [27] proposed a complex scene multi-scale ship detection model based on YOLOv7, named CSD-YOLO. This algorithm introduces a SAS-FPN module that combines dilated spatial pyramid pooling and shuffle attention, enabling the model to focus on important information while ignoring irrelevant data. This reduces feature loss for small ships and integrates feature maps of ship targets at different SAR image scales, thus improving detection accuracy and the model's ability to detect targets across multiple scales.

Researchers have proposed a series of lightweight detector algorithms based on the YOLO framework for SAR ship detection tasks. Although these algorithms perform well in target detection within SAR images, they still encounter issues of missed detections and false alarms when dealing with dense small targets in SAR images. To address the problem of low detection accuracy for dense small ship targets in complex SAR image backgrounds, this paper proposes an improved algorithm based on the YOLO11 model, named Dense Small YOLO (DS-YOLO). The main innovations are as follows:

- By utilizing distinct separation and progressive depth wise convolution operations, SPDConv [28] preserves comprehensive information across the channel dimension while effectively increasing the network's depth. This leads to more efficient feature extraction, maintaining both the richness of information and network depth, thereby enhancing the quality of feature extraction and network performance.

- The CSP-PPA module is a partially integrated pyramid attention mechanism that enhances the network's ability to capture image details and improve feature fusion. By combining CNNs and Transformers, this module effectively integrates both local and global information while reducing the number of parameters and computational time, thereby significantly improving the computational efficiency of the model.

- An adaptive normalized Wasserstein distance (AWNWD) loss function is proposed. By adaptively weighting the NWD [29] loss function and the CIoU loss function, this method enhances the focus on small target detection while balancing the shortcomings of the NWD loss function in medium and large target detection. As a result, it significantly improves the model's overall capability in multi-scale target detection.

The structure of this paper is organized as follows: Section 2 presents detailed explanations of the specific implementation process of our method. Section 3 discusses the dataset, evaluation metrics, and experimental setup in detail. Section 4 presents the experimental results along with a comparative analysis. Finally, Section 5 concludes the paper by summarizing our research findings and outlining potential directions for future studies.

2. Materials and Methods

2.1 The Network Structure of YOLO11

This paper presents improvements to the YOLO11 model, which is a lightweight object detection architecture. YOLO11 builds upon the success of its predecessors by incorporating new features designed to further enhance its performance. The network architecture of YOLO11 is illustrated in Fig. 1.

The model incorporates a new backbone network and an anchor-free detection head. The backbone network of YOLO11 draws inspiration from the design philosophy of YOLOv8, replacing the C2f module of YOLOv8 with the C3k2 module, thereby offering a richer gradient flow. Additionally, the detection head adopts the mainstream decoupled head structure, transitioning from an anchor-based approach to an anchor-free method, which directly predicts the targets' locations and sizes.

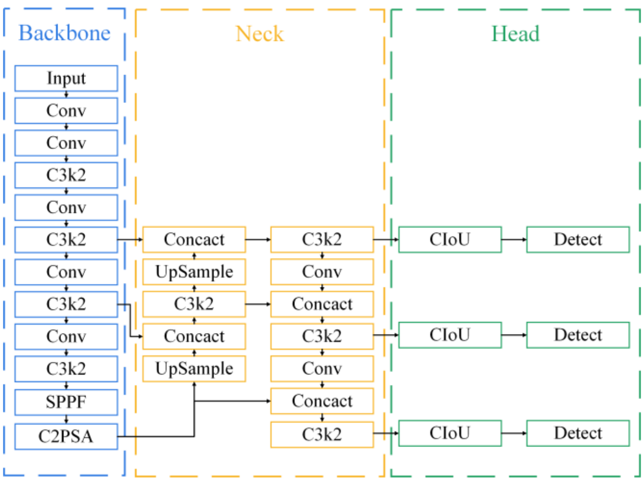


Fig. 1. Structure of YOLO11.

2.2 The Improved Network Structure

SAR images pose unique challenges for ship detection, including speckle noise, low target-background contrast, and dense small target distributions. These factors often lead to missed detections and false alarms in existing YOLO models, particularly in nearshore scenes. To address these issues, we propose DS-YOLO, an enhanced YOLO11-based model tailored for SAR ship detection. DS-YOLO replaces the original Conv module with SPDConv to preserve fine-grained features of small targets. The C3k2 module is substituted with CSP-PPA to balance local and global feature extraction, mitigating complex background interference. Additionally, a small-target detection head with AWNWD loss function is introduced to enhance adaptability across multi-scale targets, ensuring robust performance in diverse SAR scenarios. The overall architecture is shown in Fig. 2.

2.3 SPDConv Modules

In SAR target detection, small target detection remains a highly challenging problem. Due to the unique imaging mechanism of SAR, small targets in the images typically exhibit the following characteristics: firstly, small targets occupy a relatively small pixel area in the image, typically no more than 32×32 pixels; secondly, the contrast between small targets and the background is low, and they are susceptible to interference from speckle noise; and finally, statistical data shows that SAR images contain a large number of small targets, and results in significant issues such as false alarms and missed detections.

To improve the detection accuracy of small targets in SAR images and reduce missed detections, this study replaces traditional convolutional and pooling layers with the SPDConv module, as shown in Fig. 3. The SPDConv module, by employing a dense connection mechanism and a specialized feature extraction approach, is able to more effectively capture the fine details of small targets. Unlike traditional convolutional layers, the SPDConv module increases the number of feature channels at each layer. This enhancement allows for more effective extraction of key

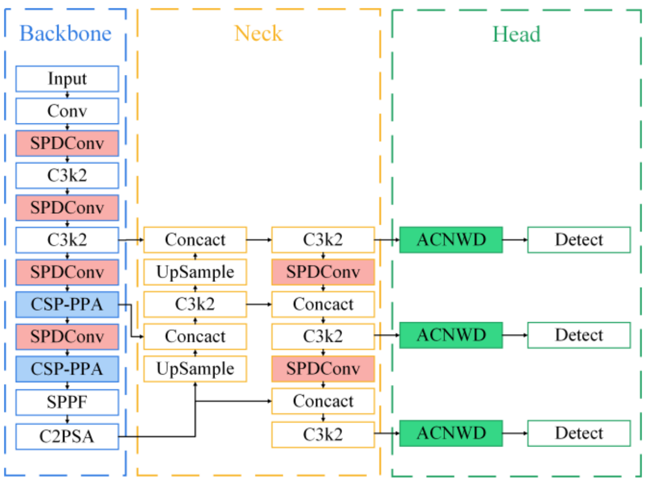


Fig. 2. Structure of DS-YOLO.

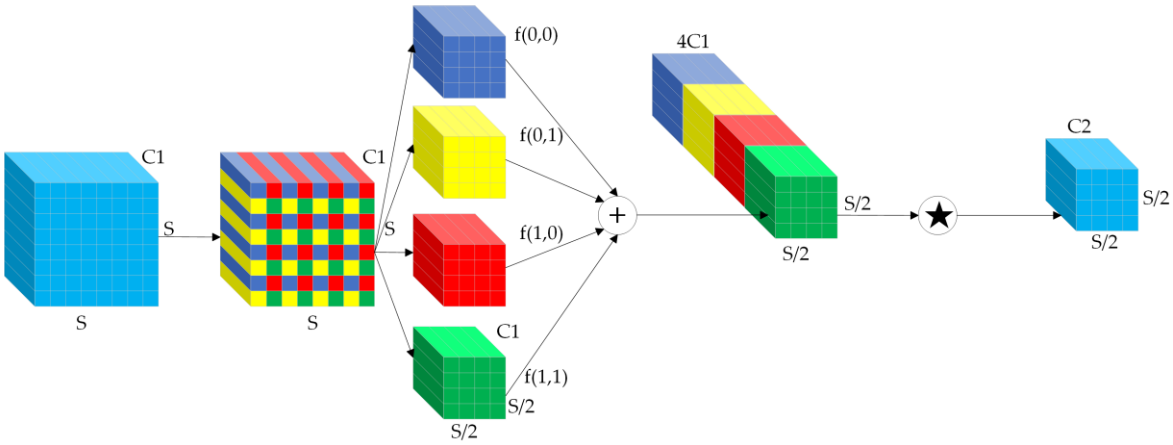


Fig. 3. The structure of SPDConv module.

features from small targets, thereby improving detection accuracy. In this way, the SPDConv module helps to reduce the feature loss and information bottleneck issues that traditional CNN architectures may encounter when dealing with small targets.

This module combines an SPD layer with a stride-free convolutional layer, completely replacing traditional convolutional strides and pooling layers. The SPD layer reduces the spatial dimensions of feature maps while preserving the integrity of information in the channel dimension, thereby avoiding information loss. This transformation is achieved by mapping each pixel or feature of the input feature map to a new channel, thus increasing the size of the channel dimension while decreasing the spatial dimension size. Considering the instance where the scaling factor is set to 2, let us posit an intermediate feature map with a spatial resolution of $S \times S$ and a channel depth of $C1$. By applying depth-wise convolution, this feature map is partitioned into a sequence of sub-feature maps. The mathematical formulation delineating the sub-feature map can be expressed as:

$$f_{scale-1,scale-1} = X[scale-1:S:scale-1:S:scale]. \quad (1)$$

Through the processing shown in (1), four sub-feature maps are obtained, namely $f_{0,0}$, $f_{0,1}$, $f_{1,0}$, and $f_{1,1}$, each with a shape of ($S/2$, $S/2$, $C1$). Then, the module concatenates these sub-feature maps along the channel dimension to obtain the feature map X' ($S/2$, $S/2$, $4C1$), whose spatial dimensions are reduced by a factor of 4. The feature map X' is further processed through $C2$ filters and is transformed into the feature map X'' ($S/2$, $S/2$, $C2$).

Figure 4 shows the heatmap of ship targets extracted using Conv and SPDConv. According to the visualization results of Hires-CAM [30], SPDConv performs better in capturing and processing the features of small targets, allowing for more accurate localization and identification of these targets. In Fig. 4(a), the targets that were not detected by traditional convolution are highlighted with red boxes, while in Fig. 4(b), the additional targets identified by SPDConv are marked with green boxes. By comparison, it can be concluded that SPDConv demonstrates stronger robustness compared to traditional convolution when addressing the challenges of dense small target detection in SAR images.

2.4 CSP-PPA Modules

In the field of computer vision, the Transformer architecture has earned widespread recognition and acclaim for its powerful ability to capture global features. However, while SPDConv achieves high precision, it inevitably reduces recall. To improve and balance both metrics while reducing false detections, we propose leveraging Transformers. Nevertheless, the high computational complexity of Transformer models presents significant challenges when applied directly to large-scale data channels, often resulting in substantial resource consumption and reduced efficiency. To address the issues of excessive parameters and high computational complexity, this paper introduces a hybrid architecture module, CSP-PPA, to replace the C3k2 module in YOLO11.

The proposed module improves the BottleNeck by dividing the input feature map into two complementary parts, which are processed separately by CNN and Transformer [31]. By combining the efficiency of CNN in local feature extraction with the advantages of the Transformer in modeling global dependencies, the new network structure significantly reduces the computational burden while main-

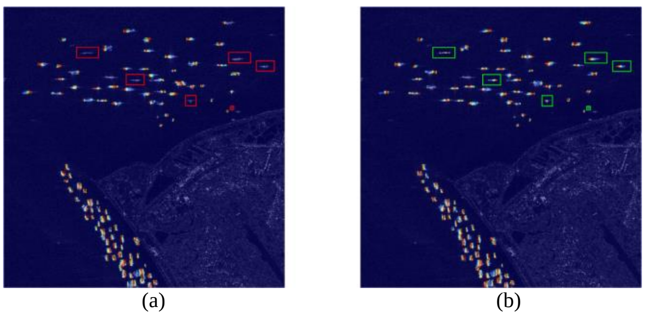


Fig. 4. The heatmap results: (a) Results of YOLO11n; (b) Results of YOLO11n with SPDConv.

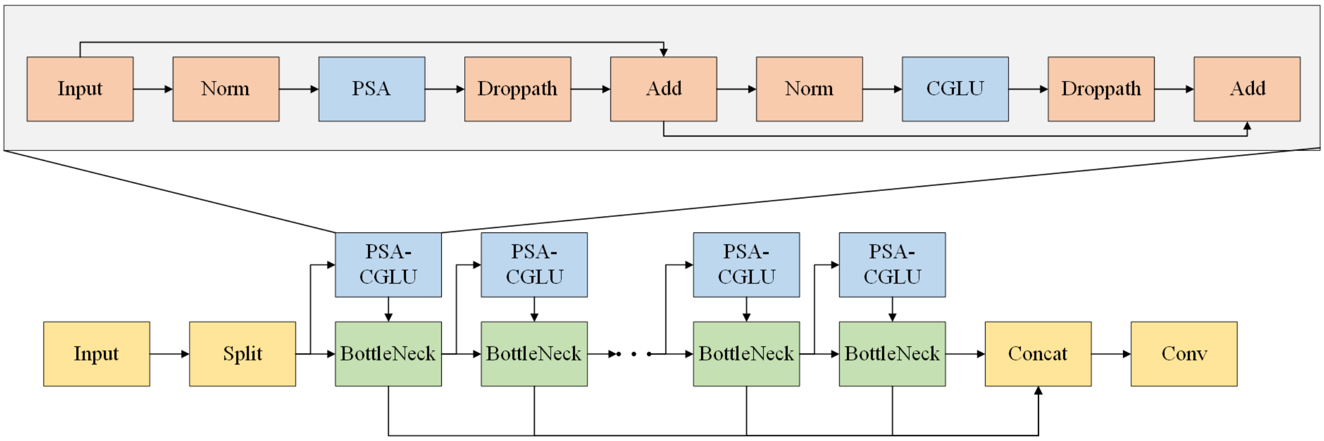


Fig. 5. The structure of CSP-PPA module.

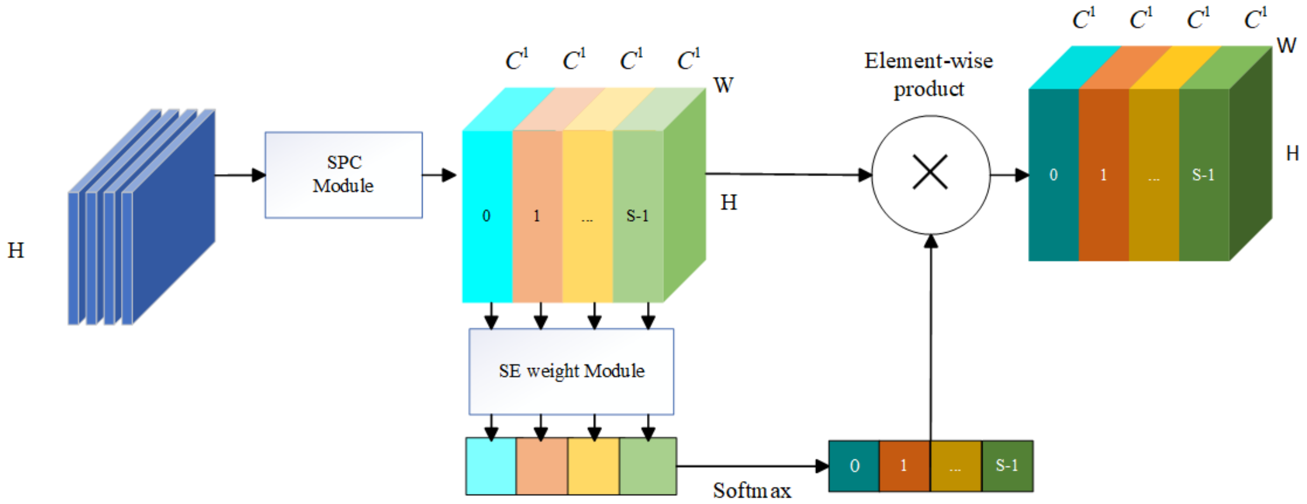


Fig. 6. Details of the PSA module.

taining efficient feature learning capabilities, as illustrated in Fig. 5.

Through this parallel and complementary processing approach, the proposed hybrid architecture not only retains the powerful global feature extraction capabilities of the Transformer but also significantly reduces the overall computational cost by incorporating CNN. This design effectively enhances the detection of ships in densely berthed near-shore areas, achieving a dual improvement in efficiency and performance, while ensuring the effective utilization of both global and local features.

The PSA module, as illustrated in Fig. 6, is implemented through the following four steps:

First, the SPC module, as shown in Fig. 7, is used to segment the channels, followed by multi-scale feature extraction of spatial information for each channel feature map.

Second, the SEWeight module extracts channel attention from feature maps of different scales, generating channel attention vectors for each scale.

Third, Softmax is applied to recalibrate the multi-scale channel attention vectors, producing new attention weights after multi-scale channel interaction.

Fourth, the recalibrated weights are element-wise multiplied with the corresponding feature maps, outputting feature maps with multi-scale attention weighting, thereby enhancing the representation of multi-scale information.

The traditional Feed Forward Networks (FFN) are replaced by the CGLU from TransNeXt [32], as shown in Fig. 8. The Multi-Head Self-Attention mechanism extracts global features, while the gated linear unit enhances nonlinear feature expression. Compared to the traditional FFN, CGLU demonstrates superior performance. Additionally, a factor is designed to control the number of input channels required by the attention mechanism, ensuring the manageability of parameters.

To validate the time efficiency of our method, this paper conducted comparative experiments using the HRSID dataset as an example, with the time unit in hours. The comparison methods include the C3k2 module of YOLO11, C3k2-PA with the Pyramid Attention mechanism (PA), CSP-PA, and our method CSP-PPA. The results shown in Tab. 1 demonstrate that our approach achieves superior temporal efficiency compared to directly integrating the PA attention mechanism into the original C3k2 model. The relatively lower time consumption of CSP-PA

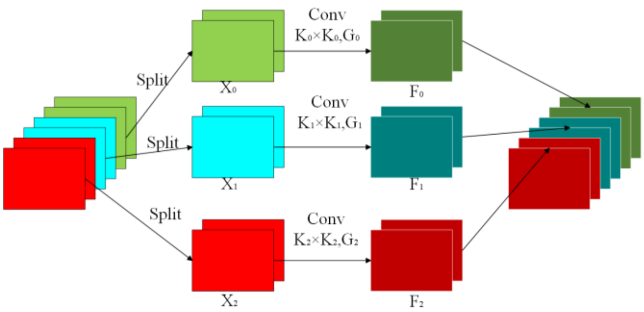


Fig. 7. Details of the SPC module.

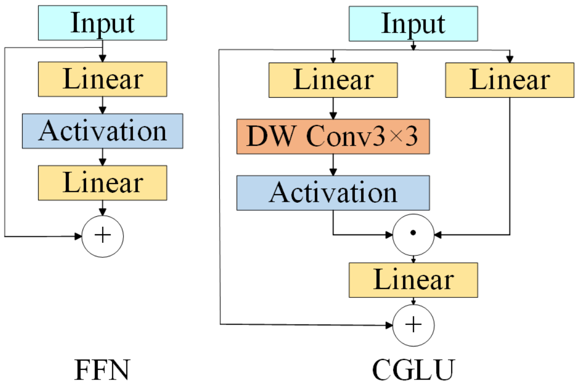


Fig. 8. Details of the CGLU module.

Method	Time	Parameters
C3k2	2.599 h	2.58 M
C3k2-PA	3.103 h	2.62 M
CSP-PA	2.685 h	2.49 M
CSP-PPA	2.7 h	2.46 M

Tab. 1. Module comparison experiment.

versus CSP-PPA may be attributed to the fact that a singular attention mechanism is more suited for parallelization, allowing for more effective utilization of hardware acceleration capabilities. In contrast, the hybridization of CNN with attention mechanisms, although increasing the complexity of parallel computation and consequently impacting overall efficiency, also results in a notable reduction in the parameter count of the module.

2.5 AWNWD Loss Function

In SAR images, the size of ship targets varies greatly, including both densely packed small vessels and large ships. Existing loss functions have limitations when dealing with such multi-scale targets. Traditional loss functions struggle to effectively handle dense small targets, while some loss functions specifically optimized for small targets may experience significant performance degradation when facing multi-scale targets detection. To address this issue, this paper proposes the AWNWD loss function, which combines the advantage of CIoU loss with the NWD loss. The AWNWD loss function introduces adaptive parameters a_1 and a_2, which automatically adjust the weights of the NWD and CIoU losses based on the size of the targets. This design enables the AWNWD loss function to flexibly handle the detection requirements of targets of different sizes, thereby maintaining the performance of small target detection while minimizing the impact on medium or large targets. The AWNWD loss function can be expressed as:

$$L_{\mathrm{AWNWD}} = a_1 L_{\mathrm{CIoU}} + a_2 L_{\mathrm{NWD}}, \tag{2}$$

$$a_1 = \frac{\mathrm{CIoU}}{\mathrm{CIoU} + \mathrm{NWD}}, \tag{3}$$

$$a_2 = 1 - a_1. \tag{4}$$

The NWD loss function is a loss function designed to enhance the accuracy and robustness of small target detection. This loss function models the target's bounding box as a two-dimensional Gaussian distribution and calculates the similarity of the target's bounding box through the corresponding Gaussian distributions, thereby making it less sensitive to the size of the predicted boxes. The NWD loss function can be expressed as:

$$L_{\mathrm{NWD}} = 1 - \mathrm{NWD}\left(N_{\mathrm{p}}, N_{\mathrm{g}}\right) \tag{5}$$

where the $\mathrm{NWD}(N_{\mathrm{p}}, N_{\mathrm{g}})$ denotes measurement between the predicted box and the ground truth box, and its calculation method can be expressed as:

$$\mathrm{NWD}\left(N_{\mathrm{p}}, N_{\mathrm{g}}\right) = \exp\left(-\frac{\sqrt{W_2^2\left(N_{\mathrm{p}}, N_{\mathrm{g}}\right)}}{C}\right). \tag{6}$$

The parameter C is a hyperparameter related to the dataset, which is determined through network training. $W_2^2(N_{\mathrm{p}}, N_{\mathrm{g}})$ is normalized in the exponential form of the Wasserstein distance, which can be expressed as:

$$W_2^2\left(N_{\mathrm{p}}, N_{\mathrm{g}}\right) \left\| \left(\left[cx_{\mathrm{p}}, cy_{\mathrm{p}}, \frac{w_{\mathrm{p}}}{2}, \frac{h_{\mathrm{p}}}{2} \right]^{\mathrm{T}}, \left[cx_{\mathrm{g}}, cy_{\mathrm{g}}, \frac{w_{\mathrm{g}}}{2}, \frac{h_{\mathrm{g}}}{2} \right]^{\mathrm{T}} \right) \right\| \tag{7}$$

where cx_{p}, cy_{p} and cx_{g}, cy_{g} represent the coordinates of the center points of the predicted box and the ground truth box, respectively, while w_{p}, h_{p}, w_{g}, h_{g} represent the height and width of the predicted box and the ground truth box, respectively.

The Wasserstein distance between the predicted box and the ground truth box is calculated based on their Gaussian distributions. N_{p} ($\boldsymbol{\mu}_{\mathrm{p}}$, $\sum_{\mathrm{p}}$), N_{g} ($\boldsymbol{\mu}_{\mathrm{g}}$, $\sum_{\mathrm{g}}$) represents the two-dimensional Gaussian distributions of the predicted box and the ground truth box, which can be expressed as:

$$\Sigma_{\mathrm{p}} = \begin{bmatrix} \frac{w_{\mathrm{p}}^2}{4} & 0 \\ 0 & \frac{h_{\mathrm{p}}^2}{4} \end{bmatrix}, \Sigma_{\mathrm{g}} = \begin{bmatrix} \frac{w_{\mathrm{g}}^2}{4} & 0 \\ 0 & \frac{h_{\mathrm{g}}^2}{4} \end{bmatrix}, \tag{8}$$

$$\mu_{\mathrm{p}} = \begin{bmatrix} cx_{\mathrm{p}} \\ cy_{\mathrm{p}} \end{bmatrix}, \mu_{\mathrm{g}} = \begin{bmatrix} cx_{\mathrm{g}} \\ cy_{\mathrm{g}} \end{bmatrix}. \tag{9}$$

And the loss function CIoU can be expressed as:

$$L_{\mathrm{CIoU}} = 1 - \mathrm{IoU} + \frac{\rho^2\left(C_{\mathrm{A}} + C_{\mathrm{B}}\right)}{d^2} + \alpha v. \tag{10}$$

IoU is the Intersection over Union between the predicted box A and the ground truth box B. $\rho(C_{\mathrm{A}}+C_{\mathrm{B}})$ is the Euclidean distance between the center point of the predicted box C_{A} and the center point of the ground truth box C_{B}. d^2 is the diagonal length of the smallest enclosing rectangle covering the predicted box and the ground truth box. v is the aspect ratio consistency term, used to measure the difference in aspect ratio between the predicted box and the ground truth box. Its calculation formula can be expressed as:

$$v = \frac{4}{\pi^2}\left(\arctan\left(\frac{W_{\mathrm{A}}}{h_{\mathrm{A}}}\right) - \arctan\left(\frac{W_{\mathrm{B}}}{h_{\mathrm{B}}}\right)\right)^2. \tag{11}$$

W_{A} and h_{A} represent the width and height of the predicted box, respectively. W_{B} and h_{B} represent the width and height of the ground truth box, respectively. α is a weighting coefficient used to balance the influence of the center point distance and the aspect ratio consistency term. Its calculation formula can be expressed as:

$$\alpha = \frac{v}{(1 - \mathrm{IoU}) + v}. \tag{12}$$

3. Experiments

This section begins with a detailed description of the experimental environment. It then systematically elaborates on the datasets used in the experiments, including the original dataset and the dense small-target dataset constructed based on it, and introduces the evaluation metrics and related experimental details. On this basis, the paper conducts comparative experiments and analyses between the proposed method and traditional SAR target detection algorithms as well as deep learning-based SAR target detection algorithms, thoroughly validating the effectiveness of the proposed method. To further validate the innovativeness of the method, the paper also carries out ablation experiments and performs visualization, analysis, and discussion of feature maps, thereby comprehensively demonstrating the advantages and contributions of the proposed method.

3.1 Environment

The platform used for the experiments is the Ubuntu system, and the code is written in PyTorch. The specific experimental environment is shown in Tab. 2.

Parameter	Configuration
Operating System	Linux
GPU	NVIDIA GeForce RTX 3090
Python Environment	Python 3.9.0
Learning Framework	PyTorch 2.3.0
CUDA Version	CUDA 12.1

Tab. 2. Detailed information of environment.

3.2 Datasets

To evaluate the performance of our method, this paper uses two publicly available datasets, SAR-Ships dataset [33] and HRSID [34], to validate the effectiveness of the network. The SAR-Ships dataset, constructed by Wang et al., contains 43 819 images from three different SAR satellites, with a total of 59 535 ships. This dataset includes ship targets with different polarization modes, resolutions, locations and scenes, and sizes. The HRSID contains 5 604 high-resolution SAR images and 16 951 ship instances, including SAR images with different resolutions, polarizations, sea conditions, sea areas, and coastal ports. During the experiment, all data are randomly divided into training, validation, and test sets in a ratio of 7:1:2. For a fair comparison, all training is conducted with default data for 300 epochs.

In the experiments, we adopted multiple metrics to comprehensively evaluate the performance of the algorithm. These metrics include Precision (P), Recall (R), mean Average Precision (mAP), mAP50-95 and F1-Score (F1):

$$\mathrm{Precision}\,(P) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{13}$$

$$\mathrm{Recall}\,(R) = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{14}$$

$$\mathrm{mAP}_c = \frac{1}{c}\sum_{c=1}^{c} \mathrm{AP}_c, \tag{15}$$

$$\mathrm{AP}_c = \frac{1}{|\mathrm{IoU}|}\sum_{\mathrm{IoU}}^{num} \mathrm{AP}_{\mathrm{IoU},c}, \tag{16}$$

$$F1 = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{17}$$

where TP (True Positives) denotes instances correctly identified as positive by the model, TN (True Negatives) denotes instances correctly identified as negative, FP (False Positives) denotes instances incorrectly identified as positive, FN (False Negatives) denotes instances incorrectly identified as negative. AP_c represents the average precision for class qc, where c represents the number of sample categories. Since there is only one category in this paper, c equals

1. 'num' represents the number of detection samples. The IoU threshold ranges from 0.5 to 0.95 with a step size of 0.05, resulting in 10 different IoU thresholds.

4. Results

4.1 Ablation Experiment

To demonstrate the detection performance of DS-YOLO, this paper conducted ablation experiments on the constructed HRSID and SAR-Ships dataset. As shown in Tab. 3, the results highlight the effectiveness of our proposed method. These experiments systematically evaluate the contributions of individual components, such as the SPDConv module, the CSP-PPA module, and the AWNWD loss function. The ablation studies provide a comprehensive analysis of how each modification improves the overall detection accuracy, particularly in handling small targets, complex nearshore backgrounds, and varying target sizes. The results underscore the robustness and adaptability of DS-YOLO in diverse SAR ship detection scenarios.

On the HRSID dataset, our method attained an mAP50-95 score of 69.9%, outperforming the baseline score of 68.3%, which further validates its superior performance in object detection tasks. It also achieved an F1 score of 89.7%, compared to the baseline's 88.1%. On the SAR-Ships dataset, our method achieved an mAP50-95 score of 68.6%, surpassing the baseline score of 67.8%. Additionally, it obtained an F1 score of 94.6%, compared to the baseline's 94.3%. Through the epoch accuracy comparison chart, as shown in Fig. 9 and Fig. 10, our method not only converges faster during training but also demonstrates higher accuracy, particularly in the early stages of training, where it quickly widens the gap with the baseline. Overall, these results fully showcase the robust performance of DS-YOLO on the HRSID and SAR-Ships datasets, providing a reliable solution for SAR ship detection.

4.2 Ablation Experiment on Small-Target Datasets

To further validate the effectiveness of the proposed method in detecting densely distributed small targets, we conducted additional filtering on the dataset. Traditional remote sensing datasets for dense small targets typically contain dozens of targets per image, while SAR datasets have relatively fewer images that meet this criterion. Therefore, we adopted the following filtering strategy: selecting samples with 3 or more small targets per image to construct a smaller-scale dataset that better aligns with the characteristics of dense distributions. The distribution of targets of different sizes across the datasets is shown in Tab. 4.

The small-target HRSID encompasses a total of 663 images, which include 36 large-scale targets, 1 053 medium-scale targets and 6 735 small-scale targets. In contrast, the small-target SAR-Ships dataset is composed of 1 752 images, with no large-scale targets present, featuring 266 medium-scale targets and 7 562 small-scale targets. This screening procedure has enabled the datasets to concentrate more on the detection of small-target targets. We have subjected the proposed detection method to stringent testing in terms of detecting small-target targets with dense distributions, thereby fully demonstrating its satisfactory effectiveness in complex scenarios. The ablation experiments are shown in Tab. 5.

On the small-target HRSID, our method achieved an mAP50 of 86.2%, which is 6.9 percentage points higher than the baseline model's 79.3%. Additionally, the mAP50-95 reached 50.8%, representing a 4.2 percentage point improvement over the baseline's 46.6%. In terms of P, our method achieved 85.4%, surpassing the baseline's 76.9% by 8.5 percentage points, while the R value of 73.3% also exceeded the baseline's 72.7% by 0.6 percentage points. The F1 score of our method was 78.9%, outperforming the baseline's 74.7% by 4.2 percentage points. On the small-target SAR-Ships dataset, although our method's P was slight-

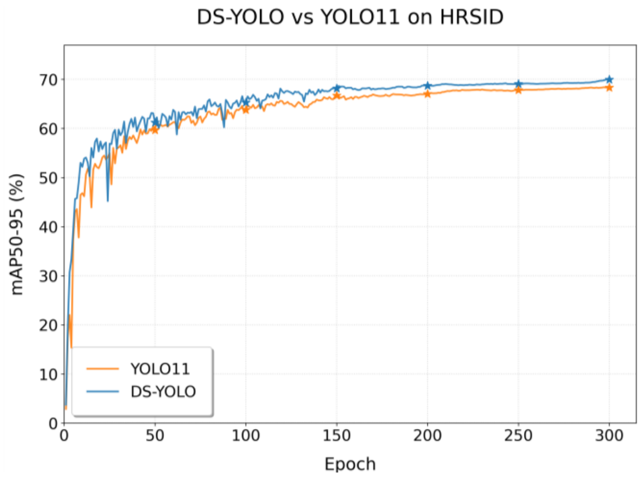


Fig. 9. Accuracy vs. epoch curve for DS-YOLO and YOLO11 on HRSID.

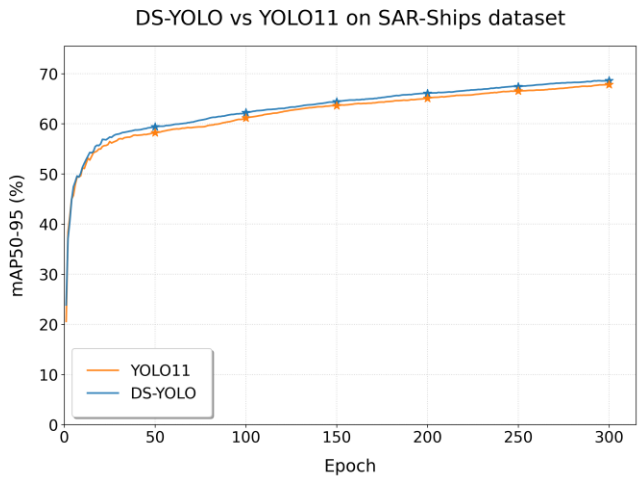


Fig. 10. Accuracy vs. epoch curve for DS-YOLO and YOLO11 on SAR-ships dataset.

Dataset	SPDConv	CSP-PPA	ACNWD	Precision	Recall	F1	mAP50	mAP50-95
HRSID	×	×	×	92.2%	84.4%	88.1%	91.8%	68.3%
	√	×	×	93.8%	84.1%	88.7%	92.7%	68.7%
	×	√	×	91.4%	84.5%	87.8%	92.2%	69.2%
	×	×	√	92.4%	84.0%	88.0%	91.8%	68.6%
	√	√	√	93.1%	**86.6%**	**89.7%**	**93.6%**	**69.9%**
SAR-Ships dataset	×	×	×	94.4%	94.2%	94.3%	97.6%	67.8%
	√	×	×	95.1%	94.1%	94.6%	97.6%	68.1%
	×	√	×	94.8%	94.3%	94.5%	97.7%	68.4%
	×	×	√	94.9%	93.5%	94.2%	97.6%	68.4%
	√	√	√	**95.5%**	93.8%	**94.6%**	**97.8%**	**68.6%**

Tab. 3. Ablation experiments.

Dataset	small	medium	large	all
HRSID	9 242	7 388	321	16 951
small-target of HRSID	6 735	1 053	36	7 824
SAR-Ships dataset	28 758	21 963	164	50 885
small-target of SAR-Ships dataset	7 562	266	0	7 828

Tab. 4. The number of targets of different sizes across various datasets.

Dataset	SPDConv	CSP-PPA	ACNWD	Precision	Recall	F1	mAP50	mAP50-95
HRSID	×	×	×	76.9%	72.7%	74.7%	79.3%	46.6%
	√	×	×	82.4%	71.0%	76.3%	81.2%	47.2%
	×	√	×	80.5%	72.5%	76.3%	80.4%	47.8%
	×	×	√	87.8%	69.7%	77.7%	82.4%	52.8%
	√	√	√	85.4%	**73.3%**	**78.9%**	**86.2%**	50.8%
SAR-Ships dataset	×	×	×	93.2%	91.0%	92.1%	95.9%	59.2%
	√	×	×	92.5%	90.2%	91.3%	95.9%	59.1%
	×	√	×	93.1%	91.0%	92.0%	95.6%	59.2%
	×	×	√	94.2%	89.6%	91.8%	95.7%	59.7%
	√	√	√	92.0%	**91.3%**	91.6%	**95.9%**	**60.4%**

Tab. 5. Ablation experiments on the small-target datasets.

ly lower than the baseline model, the mAP50-95 reached 60.4%, which is 1.2 percentage points higher than the baseline's 59.2%. These results demonstrate that our method can more stably detect densely distributed small

As shown in Fig. 11 and Fig. 12, our proposed method achieves higher accuracy compared to YOLO11. However, the performance curves exhibit greater fluctuations than those in Fig. 9 and Fig. 10, which is attributed to the reduced size and increased difficulty of the dataset used in this experiment. In this study, the proposed DS-YOLO model selected validation set images from the SAR-Ships dataset and the HRSID, and utilized visualization techniques to present the detection results of densely distributed small targets, as shown in Fig. 13 and Fig. 14. In these images, blue regions denote false positives, while red regions denote missed detections. Through experimental comparisons, the proposed model in this study exhibited significant advantages in reducing missed detections. Specifically, Figure 13 (second row) and Figure 14 (first row) illustrate the detection results of tightly packed ships in offshore scenarios. These results indicate that, while the original model was prone to missed detections when dealing with such densely distributed targets, the proposed model in this study was able to effectively identify and reduce these missed detections. Moreover, Figure 13 (first row) and Figure 14 (second row) present the detection results of targets in nearshore scenarios. These results demonstrate that the proposed model was able to accurately detect objects in nearshore scenes, where missed detections often occurred due to the tight arrangement of ships in these scenarios. This suggests that the proposed model not only enhances the detection accuracy but also improves the robustness in complex maritime environments.

To further validate the effectiveness of the proposed DS-YOLO improvements, a heatmap visualization analysis was conducted on the model, as shown in Fig. 15 and Fig. 16. In Fig. 15(b) and Fig. 16(b), red circles mark the regions of false positives and missed detections in the original model. In contrast, Figure 15(c) and Figure 16(c) highlight with green circles the targets that were not detected by the original model but were successfully identified by the proposed method, indicating a significant enhancement in detection capability. Additionally, yellow circles in Fig. 15(c) and Fig. 16(c) denote the false positives in the original model that were corrected and not erroneously detected by the proposed method, further demonstrating the effectiveness of the proposed improvements in reducing false alarms. Through the heatmap visualization analysis, it is evident that the proposed improvements excel in handling densely distributed small targets. Compared to the original model, the improvements proposed in this study for DS-YOLO are better at focusing on densely distributed small targets, especially in complex coastal backgrounds, effectively reducing false positives and missed detections.

4.3 Comparative Experiments

To objectively evaluate the performance of the improved algorithm, we compared DS-YOLO with other popular algorithms under the same conditions, and the results are shown in Tab. 6. On the two datasets, our model outperforms the two-stage Faster-RCNN and the one-stage models FCOS [35], RetinaNet, and LSKNet [36] comprehensively. When compared with the benchmark models of the YOLO series [37–39] and the latest SOTA model Hyper-YOLO [40], it achieves the highest score in the critical metric mAP50-95 model Hyper-YOLO [40], it achieves the highest score in the critical metric mAP50-95.

Through comparative analysis, the DS-YOLO model demonstrates significantly superior performance on both the HRSID and SAR-Ships datasets compared to other mainstream algorithms, including the latest version, YOLO11. On the HRSID, DS-YOLO achieves an mAP50 of 93.6% and an mAP50-95 of 69.9%, which are 1.8% and 1.6% higher than YOLO11, and 1.2% and 2.7% higher than the latest Hyper-YOLO, respectively. On the SAR-Ships dataset, DS-YOLO's mAP50 and mAP50-95 reach 97.8% and 68.6%, respectively, which are 0.2% and 0.8% higher than YOLO11, and 0.1% and 0.1% higher than Hyper-YOLO.

These results demonstrate that the DS-YOLO model achieves higher accuracy and robustness in SAR ship detection tasks, delivering superior detection performance while maintaining a lightweight architecture with only 4.5M parameters—slightly larger than some YOLO-series models but significantly more compact compared to traditional detectors such as Faster R-CNN and RetinaNet.

Moreover, the DS-YOLO model not only performs exceptionally well in the mAP50 metric but also exhibits remarkable performance in the more challenging mAP50-95 metric, thereby validating its advancement and practicality in the field of object detection.

4.4 Comparative Experiments on Small-Target Datasets

To further validate our method, we also conducted experiments on our small-target datasets, as shown in Tab. 7. Our method achieved the highest scores on the critical datasets. Upon comparative analysis, the DS-YOLO model outperforms other mainstream algorithms on both the small-target datasets of HRSID and SAR-Ships dataset. On the small-target HRSID, DS-YOLO achieves an mAP50 of 86.2% and an mAP50-95 of 50.8%, which are 4.1% and 4.7% higher, respectively, than the corresponding metrics of the latest Hyper-YOLO. On the small-target SAR-Ships dataset, DS-YOLO's mAP50 and mAP50-95 are 95.9% and 60.4%, respectively, continuing to lead other algorithms. Although it is slightly lower than Hyper-YOLO by 0.3% in mAP50 on this dataset, it is 0.7% higher in the more challenging mAP50-95 metric.

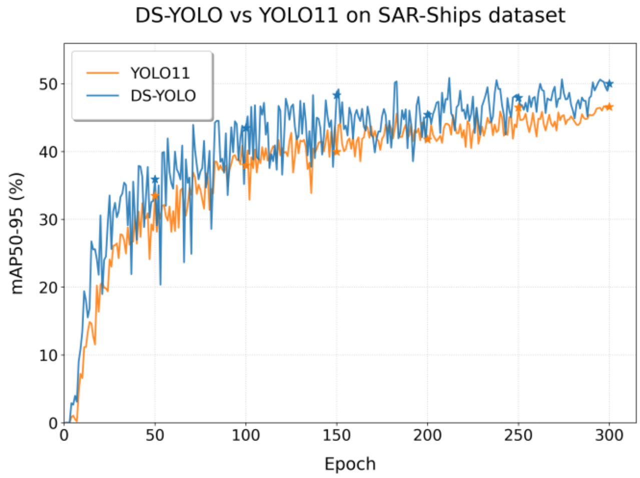


Fig. 11. Accuracy vs. epoch curve for DS-YOLO and YOLO11 on HRSID.

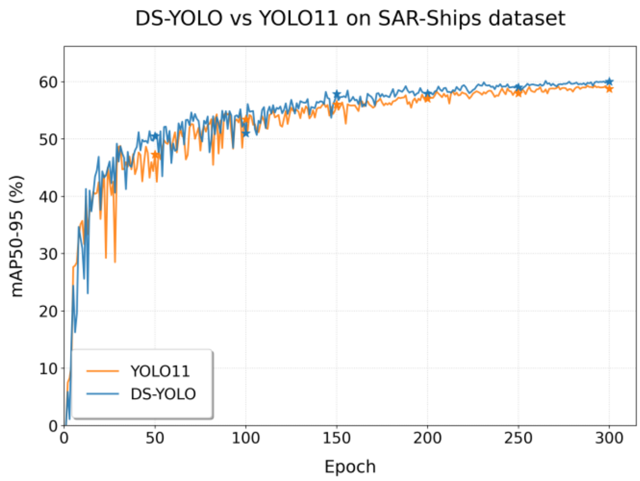
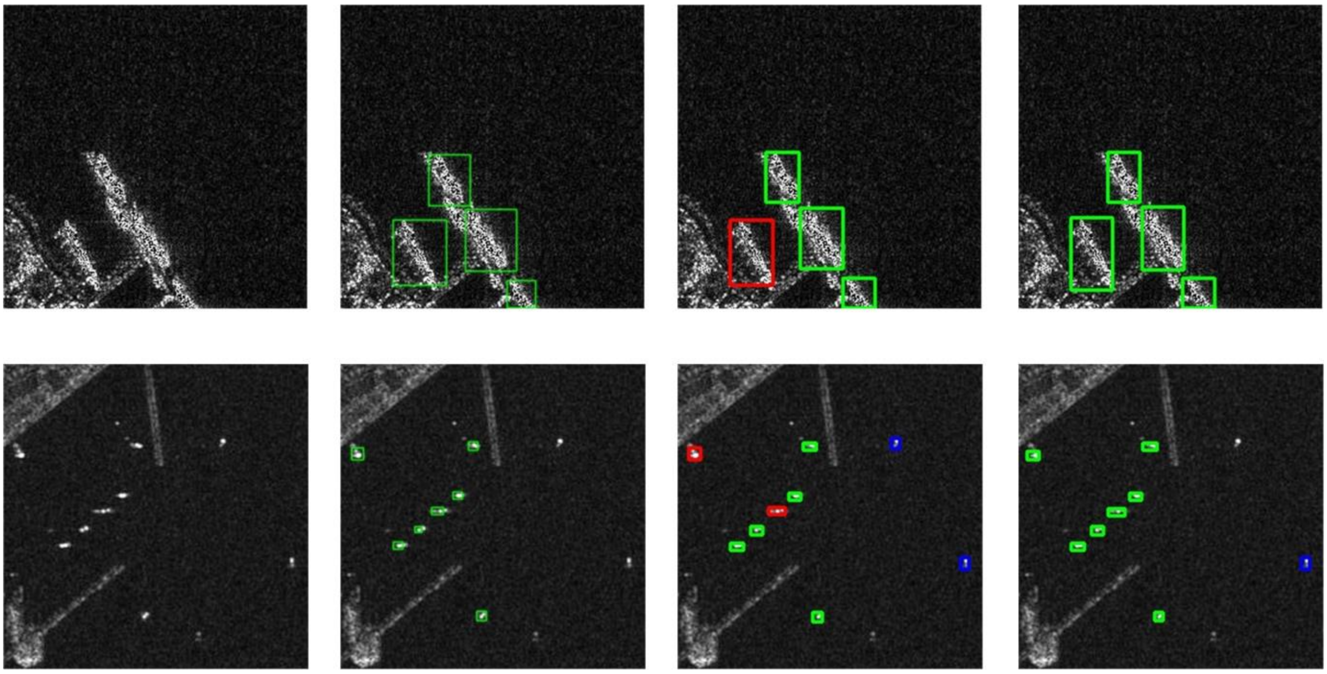


Fig. 12. Accuracy vs. epoch curve for DS-YOLO and YOLO11 on SAR-Ships dataset.

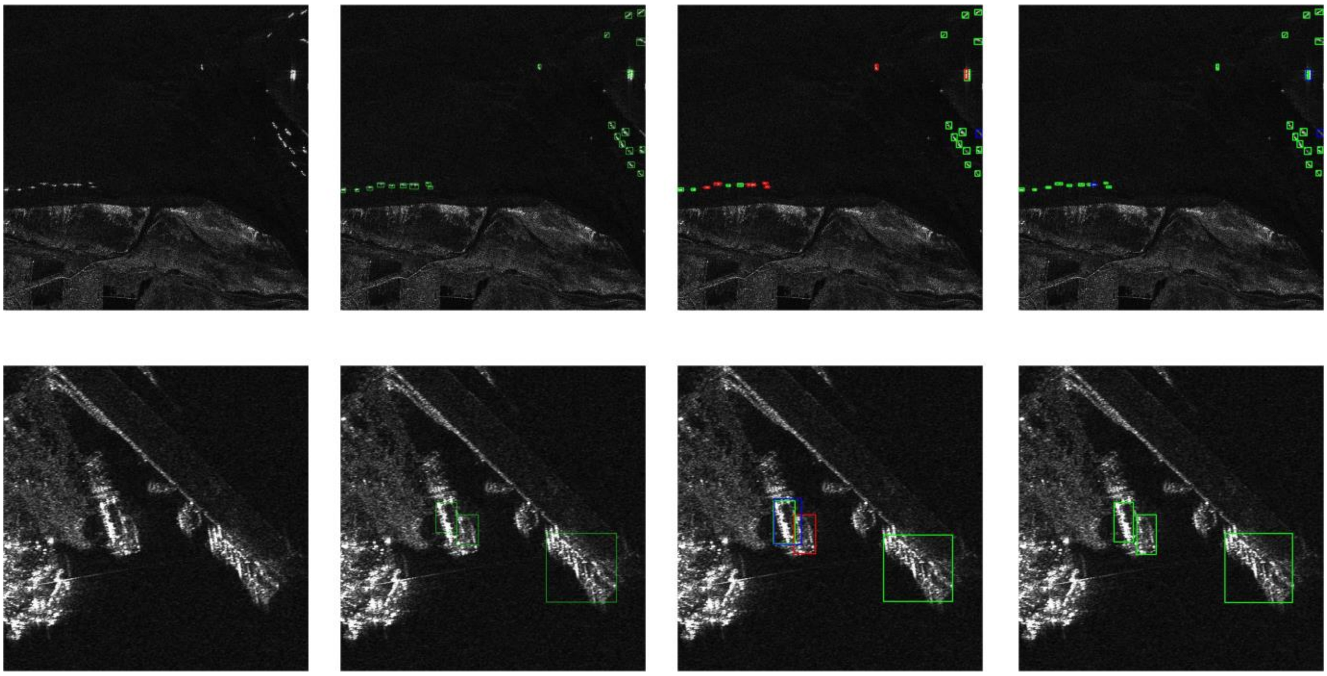
5. Conclusions

This paper proposes an improved ship detection algorithm based on YOLO11, named DS-YOLO, with a focus on enhancing the accuracy of detecting small targets, especially dense small targets. The SPDConv is introduced to better extract features of small targets. The CSP-PPA module is proposed, which combines Transformer and CNN to demonstrate precision advantages and also alleviates time consumption to some extent. A novel AWNWD loss function is proposed. By adaptively adjusting weights, it combines the strengths of two loss functions while amplifying their respective advantages. Experimental results demonstrate that, compared to current mainstream object detection algorithms, DS-YOLO achieves more accurate ship recognition on the HRSID and SAR-Ships dataset. Compared



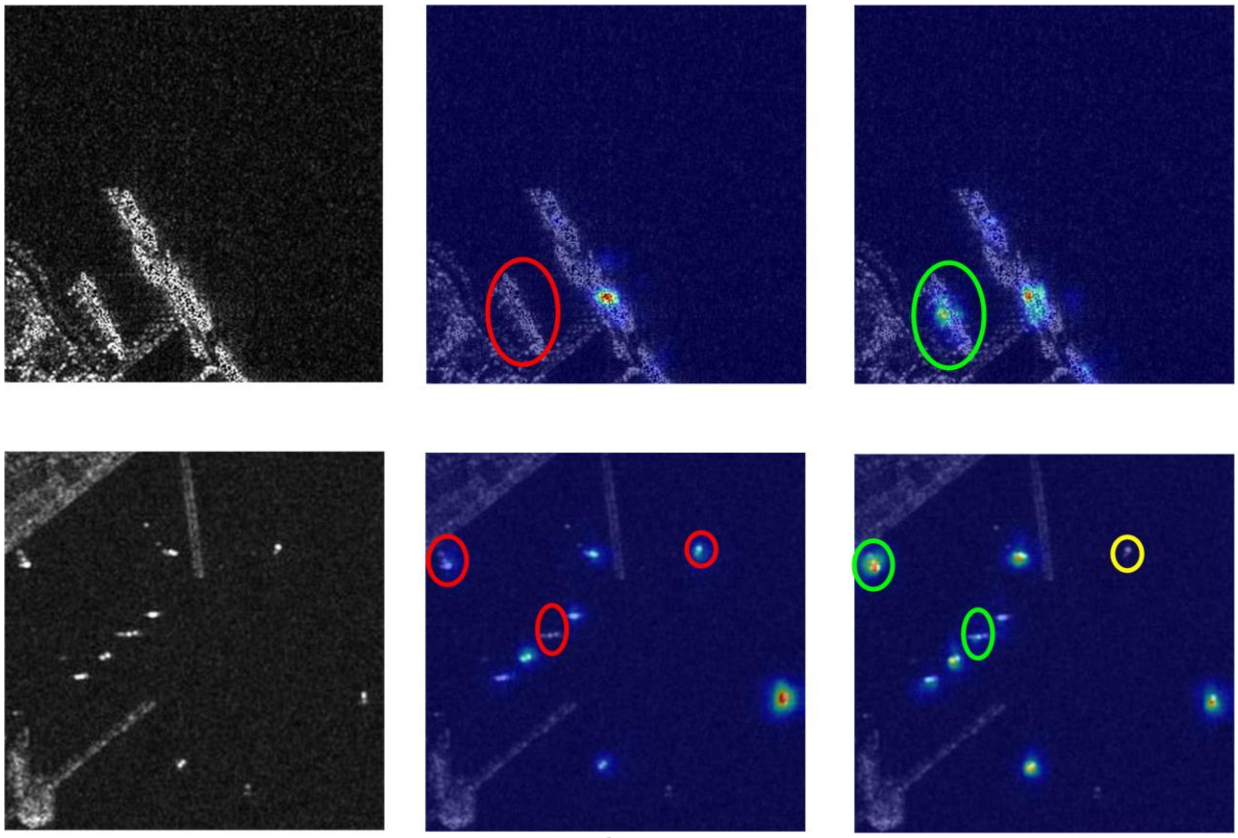
(a) (b) (c) (d)

Fig. 13. SAR ship detection results on the SAR-Ships dataset: (a) Background; (b) Ground truth; (c) Detection results of YOLO11n; (d) Detection results of DS-YOLO.



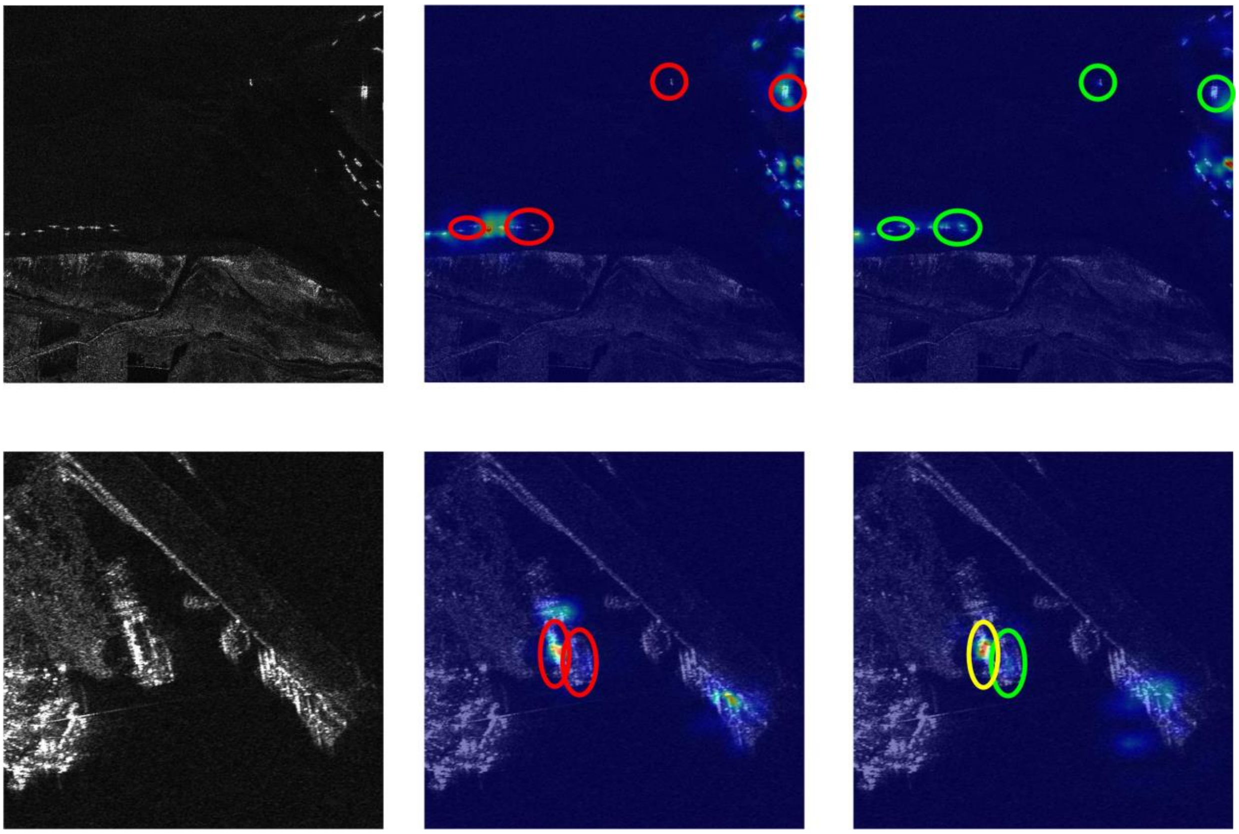
(a) (b) (c) (d)

Fig. 14. SAR ship detection results on the HRSID: (a) Background; (b) Ground truth; (c) Detection results of YOLO11n; (d) Detection results of DS-YOLO.



(a) (b) (c)

Fig. 15. The heatmap results on the SAR-Ships dataset: (a) Background; (b) Detection results of YOLO11n; (c) Detection results of DS-YOLO.



(a) (b) (c)

Fig. 16. The heatmap results on the HRSID: (a) Background; (b) Detection results of YOLO11n; (c) Detection results of DS-YOLO.

Dataset	Parameters (M)	Method	Precision	Recall	mAP50	mAP50-95
HRSID	32.1	FCOS	—	—	83.3%	56.6%
	41.4	Faster-RCNN	—	—	86.2%	63.7%
	32.5	RetinaNet	—	—	89.2%	65.5%
	19.1	LSKNet	91.5%	83.9%	91.6%	68.2%
	3.1	YOLOv7t	88.0%	75.3%	84.2%	56.1%
	3.2	YOLOv8n	88.0%	85.5%	90.4%	65.5%
	2	YOLOv9t	89.2%	80.5%	89.4%	64.6%
	2.3	YOLOv10n	90.2%	83.5%	90.7%	66.5%
	2.6	YOLO11n	92.2%	84.4%	91.8%	68.3%
	3.1	hyper-YOLO	93.2%	84.2%	92.4%	67.2%
	4.5	ours	93.1%	**86.6%**	**93.6%**	**69.9%**
SAR-Ships dataset	32.1	FCOS	—	—	96.1%	60.1%
	41.4	Faster-RCNN	—	—	95.4%	59.3%
	32.5	RetinaNet	—	—	96.9%	64.5%
	19.1	LSKNet	93.9%	93.5%	97.1%	66.7%
	3.1	YOLOv7t	93.5%	92.9%	96.3%	62.0%
	3.2	YOLOv8n	94.3%	94.6%	97.6%	68.7%
	2	YOLOv9t	93.8%	92.8%	97.3%	66.7%
	2.3	YOLOv10n	91.0%	90.4%	95.2%	61.3%
	2.6	YOLO11n	94.4%	94.2%	97.6%	67.8%
	3.1	hyper-YOLO	95.1%	94.8%	97.7%	68.5%
	4.5	ours	**95.5%**	93.8%	**97.8%**	**68.6%**

Tab. 6. Comparison of the performance metrics of different models.

Dataset	Parameters (M)	Method	Precision	Recall	mAP50	mAP50-95
HRSID	32.1	FCOS	—	—	41.7%	18.5%
	41.4	Faster-RCNN	—	—	68.8%	39.4%
	32.5	RetinaNet	—	—	68.3%	38.7%
	19.1	LSKNet	78.9%	73.7%	80.5%	47.2%
	3.1	YOLOv7t	77.0%	68.2%	74.2%	36.9%
	3.2	YOLOv8n	83.2%	74.2%	82.4%	50.2%
	2	YOLOv9t	82.5%	69.7%	79.6%	45.8%
	2.3	YOLOv10n	84.6%	73.3%	82.8%	48.7%
	2.6	YOLO11n	76.9%	72.7%	79.3%	46.6%
	3.1	hyper-YOLO	79.9%	75.6%	82.1%	46.1%
	4.5	ours	**85.4%**	73.3%	**86.2%**	**50.8%**
SAR-Ships dataset	32.1	FCOS	—	—	83.2%	38.1%
	41.4	Faster-RCNN	—	—	91.9%	49.6%
	32.5	RetinaNet	—	—	94.0%	50.8%
	19.1	LSKNet	90.7%	90.6%	95.1%	55.9%
	3.1	YOLOv7t	90.4%	90.5%	94.7%	54.9%
	3.2	YOLOv8n	90.5%	90.8%	95.0%	58.3%
	2	YOLOv9t	92.6%	89.2%	95.8%	57.8%
	2.3	YOLOv10n	89.1%	91.6%	94.9%	57.3%
	2.6	YOLO11n	93.2%	91.0%	95.9%	59.2%
	3.1	hyper-YOLO	95.0%	91.1%	96.2%	59.7%
	4.5	ours	92.0%	91.3%	95.9%	**60.4%**

Tab. 7. Comparison of the performance metrics of different models on small-target datasets.

to the baseline model, the mAP50-95 metric is improved by 1.6% and 0.8%. On the small target datasets, the mAP50-95 metric is improved by 4.2% and 1.2%, respectively. The proposed algorithm demonstrates excellent performance on small target datasets while also achieving competitive results on general target datasets. Visualization results further validate the effectiveness and robustness of the proposed DS-YOLO in complex scenarios. The advancements offered by DS-YOLO have significant implications for remote sensing and computer vision applications. By improving small target detection in SAR imagery, DS-YOLO holds promise for diverse applications, including maritime surveillance, port management, illegal fishing monitoring, and disaster response, where dense small objects are common. Future work will focus on enhancing detection accuracy for dense small targets in SAR images through advanced data augmentation techniques, leveraging diffusion models to generate diverse training samples. This approach aims to strengthen the model's feature extraction capabilities, particularly in complex scenarios involving dense target distributions. The source code for DS-YOLO is publicly available at https://github.com/ShenyiFei2023/DS-YOLO/tree/main.

Acknowledgments

This work was funded by the Huzhou city Science and Technology Plan Project (Grant 2022YZ20) and the Graduate Research and Innovation Project of the School of Information Engineering at Huzhou University (2024XGKC01).

References

[1] HUMAYUN, M. F., NASIR, F. A., BHATTI, F. A., et al. YOLO-OSD: Optimized ship detection and localization in multiresolution SAR satellite images using a hybrid data-model centric approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, vol. 17, no. 1, p. 5345–5363. DOI: 10.1109/JSTARS.2024.3365807

[2] DENG, Z., SUN, H., ZHOU, S., et al. Learning deep ship detector in SAR images from scratch. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, vol. 57, no. 1, p. 4021–4039. DOI: 10.1109/TGRS.2018.2889353

[3] WANG, Y., WANG, C., ZHANG, H., et al. Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery. *Remote Sensing*, 2019, vol. 11, no. 5, p. 1–14. DOI: 10.3390/rs11050531

[4] ZHU, M., HU, G., ZHOU, H., et al. Rapid ship detection in SAR images based on YOLOv3. In *Proceedings of the 5th International Conference on Communication, Image and Signal Processing (CCISP)*. Chengdu (China), 2020, p. 214–218. DOI: 10.1109/CCISP51026.2020.9273476

[5] REDMON, J., DIVVALA, S., GIRSHICK, R., et al. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas (USA), 2016, p. 779–788. DOI: 10.1109/CVPR.2016.91

[6] LIU, W., ANGUELOV, D., ERHAN, D., et al. SSD: Single shot multibox detector. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Amsterdam (Netherlands), 2016, p. 21 to 37. DOI: 10.1007/978-3-319-46448-0_2

[7] LIN, T., GOYAL, P., GIRSHICK, R., et al. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Venice (Italy), 2017, p. 2999–3007. DOI: 10.1109/ICCV.2017.324

[8] GIRSHICK, R., DONAHUE, J., DARRELL, T., et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Columbus (USA), 2014, p. 580–587. DOI: 10.1109/CVPR.2014.81

[9] GIRSHICK, R. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Santiago (Chile), 2015, p. 1440–1448. DOI: 10.1109/ICCV.2015.169

[10] REN, S., HE, K., GIRSHICK, R., et al. Faster R-CNN: Towards real-time object detection with region proposal networks. I*EEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, vol. 39, no. 6, p. 1137–1149. DOI: 10.1109/TPAMI.2016.2577031

[11] LV, Y., LI, M., HE, Y. A novel twin branch network based on mutual training strategy for ship detection in SAR images. *Complex Intelligent Systems*, 2024, vol. 10, p. 2387–2400. DOI: 10.1007/s40747-023-01240-y

[12] CHAI, B., NIE, X., ZHOU, Q., et al. Enhanced cascade R-CNN for multi-scale object detection in dense scenes from SAR images. *IEEE Sensors Journal*, 2024, vol. 24, no. 12, p. 20143–20153. DOI: 10.1109/JSEN.2024.3393750

[13] WANG, K., QIAO, Q., ZHANG, G., et al. Few-shot SAR target recognition based on deep kernel learning. *IEEE Access*, 2022, vol. 10, no. 1, p. 89534–89544. DOI: 10.1109/ACCESS.2022.3193773

[14] ZHOU, Z., CUI, Z., TANG, K., et al. Gaussian meta-feature balanced aggregation for few-shot synthetic aperture radar target detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2024, vol. 208, p. 89–106. DOI: 10.1016/j.isprsjprs.2024.01.003

[15] ZHANG, X., LI, Y., LI, F., et al. Ship-Go: SAR ship images inpainting via instance-to-image generative diffusion models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2024, vol. 207, no. 1, p. 203–217. DOI: 10.1016/j.isprsjprs.2023.12.002

[16] ANANDHI, D., VALLI, S. An enhanced approach to despeckle SAR images. *Radioengineering*, 2018, vol. 27, no. 3, p. 864-875. DOI: 10.13164/re.2018.0864

[17] ZHANG, C., GAO, G., LIU, J., et al. Oriented ship detection based on soft thresholding and context information in SAR images of complex scenes. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, vol. 62, p. 1–15. DOI: 10.1109/TGRS.2023.3340891

[18] WANG, C., GUO, B., SONG, J., et al. A novel CFAR-based ship detection method using range-compressed data for spaceborne SAR system. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, vol. 62, no. 1, p. 1–15. DOI: 10.1109/TGRS.2024.3419893

[19] LEE, K., LEE, S., CHANG, J. A study on ship detection and classification using KOMPSAT optical and SAR images. *Ocean Science Journal*, 2024, vol. 59. DOI: 10.1007/s12601-024-00134-5

[20] LI, X., CHEN, P., YANG, J., et al. TKP-NET: A three keypoint detection network for ships using SAR imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, vol. 17, no. 1, p. 364–376. DOI: 10.1109/JSTARS.2023.3329252

[21] AMIN, B., RIAZ, M., GHAFOOR, A. Automatic image matting of Synthetic Aperture Radar target chips. *Radioengineering*, 2020, vol. 29, no. 1, p. 228–234. DOI: 10.13164/re.2020.0228

[22] REN, X., BAI, Y., LIU, G., et al. YOLO-Lite: An efficient lightweight network for SAR ship detection. *Remote Sensing*, 2023, vol. 15, no. 15, p. 1–21. DOI: 10.3390/rs15153771

[23] WU, F., HU, T., XIA, Y., et al. WDFA-YOLOX: A wavelet-driven and feature-enhanced attention YOLOX network for ship detection in SAR images. *Remote Sensing*, 2024, vol. 16, no. 10, p. 1–24. DOI: 10.3390/rs16101760

[24] FENG, Y., CHEN, J., HUANG, Z., et al. A lightweight position-enhanced anchor-free algorithm for SAR ship detection. *Remote Sensing*, 2022, vol. 14, no. 8, p. 1–19. DOI: 10.3390/rs14081908

[25] YU, J., WU, T., ZHANG, X., et al. An efficient lightweight SAR ship target detection network with improved regression loss function and enhanced feature information expression. *Sensors*, 2022, vol. 22, no. 9, p. 1–26. DOI: 10.3390/s22093447

[26] WANG, H., HAN, D., CUI, M., et al. NAS-YOLOX: A SAR ship detection using neural architecture search and multi-scale attention. *Connection Science*, 2023, vol. 35, no. 1, p. 1–32. DOI: 10.1080/09540091.2023.2257399

[27] CHEN, J., SHEN, Y., LIANG, Y., et al. YOLO-SAD: An efficient SAR aircraft detection network. *Applied Sciences-Basel*, 2024, vol. 14, no. 7, p. 1–24. DOI: 10.3390/app14073025

[28] SUNKARA, R., LUO, T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. In *Proceedings of the Machine Learning and Knowledge Discovery in Databases*. Grenoble (France), 2022, part III, p. 443–459. DOI: 10.1007/978-3-031-43320-2_27

[29] XU, C., WANG, J., YANG, W., et al. Detecting tiny objects in aerial images: A normalized Wasserstein distance and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, vol. 190, p. 79–93. DOI: 10.1016/j.isprsjprs.2022.06.002

[30] DRAELOS, R. L., CARIN, L. *HiResCAM: Explainable multi-organ multi-abnormality prediction in 3d medical images.* 11 pages. [Online] Cited 2020-11-17. Available at: https://arxiv.org/pdf/2011.08891v1

[31] VASWANI, A., SHAZEER, N., PARMAR, N., et al. Attention is all you need. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*. Long Beach (USA), 2017, p. 5998–6008. DOI: 10.48550/arXiv.1706.03762

[32] SHI, D. TransNeXt: Robust foveal visual perception for vision transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle (USA), 2024, p. 17773–17783. DOI: 10.1109/CVPR52733.2024.01683

[33] WANG, Y., WANG, C., ZHANG, H., et al. A SAR dataset of ship detection for deep learning under complex backgrounds. *Remote Sensing*, 2019, vol. 11, no. 7, p. 1–14. DOI: 10.3390/rs11070765

[34] WEI, S., ZENG, X., QU, Q., et al. A SAR dataset of ship detection for deep learning under complex backgrounds. *IEEE Access*, 2020, vol. 8, p. 120234–120254. DOI: 10.1109/ACCESS.2020.3005861

[35] TIAN, Z., SHEN, C., CHEN, H., et al. FCOS: Fully convolutional one-stage object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Seoul (South Korea), 2019, p. 9626–9635. DOI: 10.1109/ICCV.2019.00972

[36] LI, Y., LI, X., DAI, Y., et al. LSKNET: A foundation lightweight backbone for remote sensing. *International Journal of Computer Vision*, 2025, vol. 133, p. 1410–1431. DOI: 10.1007/s11263-024-02247-9

[37] WANG, C., BOCHKOVSKIY, A., LIAO, H. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver (Canada), 2023, p. 7464–7475. DOI: 10.1109/CVPR52729.2023.00721

[38] WANG, C., YEH, I., LIAO, H. M. YOLOv9: Learning what you want to learn using programmable gradient information. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Milan (Italy), 2024, p. 1–21. DOI: 10.1007/978-3-031-72751-1_1

[39] WANG, A., CHEN, H., LIU, L., et al. YOLOv10: Real-time end-to-end object detection. 21 pages. [Online] Cited 2024-5-23. Available at: https://arxiv.org/abs/2405.14458

[40] FENG, Y., HUANG, J., DU, S., et al. Hyper-YOLO: When visual object detection meets hypergraph computation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025, vol. 47, no. 4, p. 2388–2401. DOI: 10.1109/TPAMI.2024.3524377

About the Authors ...

Yifei SHEN was born in 2000. He received the B.Eng. degree in Digital Media Technology from Zhejiang Yuexiu University, Shaoxing, China, in 2022. He is currently working toward the M.E. degree in Computer Science and Technology with the Huzhou University, Huzhou, China. His research interests include SAR ship detection and machine learning.

Qiang GAO was born in 1982. He received his B.Sc. degree in 2005, M.Sc. degree in 2009, and Ph.D. degree in 2019, all from Beijing Institute of Technology. He is currently a lecturer at Huzhou University. His research interests include SAR data augmentation and SAR image interpretation.