IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 October 2024, accepted 6 November 2024, date of publication 13 November 2024, date of current version 21 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3497314

RESEARCH ARTICLE

DGSP-YOLO: A Novel High-Precision Synthetic Aperture Radar (SAR) Ship Detection Model

ZHU LEJUN, CHEN JINGLIANG, CHEN JIAYU, AND YANG HAO
School of Computer Science, Hubei University of Technology, Wuhan 430068, China
Corresponding author: Zhu Lejun (zhulejun@hbut.edu.cn)

ABSTRACT With the rapid advancement of deep learning, its application in synthetic aperture radar (SAR) ship target detection has become increasingly prevalent. However, the detection of ships in complex environments and across various scales remains a formidable challenge. This paper introduces DGSP-YOLO, a novel high-performance detection model designed to overcome these hurdles. The model integrates the SPDConv and C2fMHSA modules into the YOLOv8n baseline, significantly enhancing the feature extraction capabilities for small-scale targets. Additionally, the original convolutional blocks have been optimized with GhostConv, ensuring efficient performance and reduced parameter count. To further refine the detection process, the DySample module has been incorporated to mitigate noise interference, leading to the generation of more refined feature maps. The model also employs EIoU to bolster its capacity to process images of varying quality. Extensive experiments on the HRSID, LS-SSDD-v1.0, and SSDD datasets have been conducted to test the model's effectiveness rigorously. The results demonstrate that DGSP-YOLO outperforms other prevalent models, achieving mAP50 and mAP50:95 scores of 94% and 72.2% on the HRSID dataset, and 69% and 25.3% on the LS-SSDD-v1.0 dataset, respectively. On the SSDD dataset, the model achieved an impressive mAP50 and mAP50:95 of 99% and 75.1%, respectively. These outcomes underscore DGSP-YOLO's superior accuracy and overall performance, marking a significant advancement in SAR ship target detection.

INDEX TERMS Deep learning, synthetic aperture radar (SAR), ship target detection, YOLOv8.

I. INTRODUCTION

The application of synthetic aperture technology within the realms of radar (SAR) and sonar (SAS) [1], [2], [3], [4] has demonstrated broad applicability and advantages in various remote sensing imaging applications. As a widely applied remote sensing technology, Synthetic Aperture Radar (SAR) captures and processes signals reflected from ground targets to generate high-resolution images, with its all-weather, round-the-clock operational capability gaining widespread application and recognition. This technology holds significant value in numerous fields, including military reconnaissance [5], maritime surveillance and ship rescue [6], [7], [8], [9], [10], [11], urban planning [12], and agricultural yield estimation [13]. In particular, SAR plays a crucial role in monitoring the safety of vessels navigating vast waters in maritime surveillance. However, environmental noise and the presence of numerous small-sized targets in SAR imagery present challenges for the accuracy of ship target detection.

In recent years, a range of methods have been developed for detecting and identifying ship targets in SAR imagery, encompassing both traditional SAR detection algorithms and cutting-edge deep learning (DL) methods. Traditional SAR detection algorithms, such as the Constant False Alarm Rate (CFAR) [14] and its variants [15], hold a significant place in this domain. The CFAR algorithm establishes a detection threshold based on the statistical properties of background clutter, facilitating the identification of target pixels through grayscale value comparisons. While this algorithm strives to maintain a constant false alarm rate, its detection accuracy is closely related to the statistical characteristics of clutter. However, the efficiency of traditional detection algorithms is often hampered by their susceptibility to noise and clutter, as well as their dependence on parameter selection, which

The associate editor coordinating the review of this manuscript and approving it for publication was Xuebo Zhang.

increases the complexity of detection and poses higher demands on feature extraction, detection accuracy, and robustness.

Over the last decade, SAR-based ship detection and identification have made remarkable advances due to the rapid development of deep learning technology [16]. Deep learning detection and identification methods are broadly divided into two categories: two-stage and one-stage detection algorithms.

Two-stage methods, such as R-CNN [17], Fast R-CNN [18], and Faster R-CNN [19], rely on region proposals followed by classification and localization. While these methods excel in accuracy, they require significant computational resources, leading to slower processing speeds, which is a notable drawback in real-time applications. In contrast, one-stage methods, including SSD [20], YOLO [21], and RetinaNet [22], predict the locations and classes of all objects in a single forward pass, thus offering a speed advantage, albeit with some compromise in accuracy compared to two-stage methods. However, these methods still face challenges in dealing with small targets and complex backgrounds, especially in the context of SAR imagery, where these challenges become even more pronounced.

In SAR applications, the demand for high-speed and high-precision detection is growing, prompting researchers to continuously seek ways to enhance one-stage detection algorithms. For instance, Guo et al. [23] improved feature extraction and utilized contextual information through CenterNet++ to enhance the accuracy of small target detection, but this method may not be as effective in handling large-scale targets. Bao et al. [24] developed a complementary pre-training technique that significantly improved detection performance in SAR images, although this method requires substantial annotated data, making implementation costly in practical applications. Wang et al. [25] enhanced the capabilities of SSD through data augmentation and transfer learning, but this approach may have limited effectiveness in dealing with the specific noise and interference present in SAR images. Tang et al. [26] introduced N-YOLO, which addresses segmentation and occlusion challenges in SAR images through noise classification and complete target extraction, but this method may not be sensitive enough to rapidly moving targets. Jiang et al. [27] achieved excellent detection results by customizing YOLOv4 [28] for SAR image characteristics, but this method may lack generalization when dealing with targets under different environmental conditions. Sun et al. [29] proposed BiFA-YOLO, a dual-flow feature fusion network adept at detecting ships in any direction, though it may be computationally costly when handling high dynamic range SAR images. Wang et al. [30] designed YOLO-SD, integrating multi-scale convolution and feature transformation modules to enhance the detection of small vessels, though this method may not be precise enough when dealing with large-scale targets. Guo et al. [31] improved LMSD-YOLO, a lightweight algorithm based on YOLOv5 for multi-scale SAR ship detection, but this method may be slower when handling high-resolution images. Ren et al. [32] proposed YOLO-Lite, which optimizes network structure and feature fusion to achieve high precision and fast inference, though it may not be robust enough when dealing with complex backgrounds. Liangjun et al. [33] achieved significant success in multi-scale detection tasks by enhancing the channel and spatial attention mechanisms of YOLOv8 [34], but this method may lack sensitivity when dealing with small targets. Finally, Tang et al. [35] developed DBW-YOLO, which integrates a dynamic background weighting module to automatically adjust the weight of background features, thus making the network more focused on target areas, though this method may lack flexibility when dealing with rapidly changing environments.

These contributions reflect the dynamic evolution of SAR ship detection, with each advancement pushing the boundaries of speed, accuracy, and robustness in response to complex environments. However, ongoing research highlights the scientific community's commitment to meeting the stringent demands of SAR ship detection, ensuring the continued relevance and impact of this technology in both military and civilian applications. Despite these advances, existing algorithmic models still face substantial challenges in accurately detecting ships of varying scales, particularly those near the coastline. The primary issues stem from two main factors: firstly, SAR images typically feature high noise levels and low resolution, making it difficult for models to distinguish fine details such as the shapes and edges of background and target ships, especially in the case of small-sized ships near the coastline. Secondly, the large scale variation of ship targets in SAR images requires fine-grained features for precise detection. However, many current algorithms prioritize the detection of small targets, inevitably compromising the detection accuracy for large-scale ships.

To address these challenges, we propose a novel SAR ship detection model named DGSP-YOLO, designed to enhance the model's capacity for detecting and identifying ships across diverse backgrounds and scales. The main contributions of this research are as follows:

1. To deal with the low-resolution problem common in SAR imagery and the widespread presence of small-sized ships, we have integrated SPDConv into the YOLOv8 backbone and replaced the backbone and neck convolutions with GhostConv. This modification not only strengthens the network's ability to extract features from small targets but also refrains from substantially increasing the model's parameter count.

2. To enhance the model's capability to detect and identify multi-scale targets, we have combined C2f with MHSA Attention, enabling the application of self-attention mechanisms across features of varying scales. This integration, along with the SPPF fusion module, significantly boosts the model's detection performance for ships of different sizes.

3. We have replaced YOLOv8's conventional up-sampling layer with DySample, a module that refines edge up-sampling by dispersing a single point into multiple points and reduces model variance through a multi-stage cross-modal alignment process, thereby enhancing both the model's robustness and accuracy.

4. The introduction of EIoU aims to mitigate the sample imbalance issue in bounding box regression, steering the model to focus more on optimizing prediction frames with higher overlap with the ground truth frames.

The subsequent sections of this paper are organized as follows: Section II elaborates on the intricacies of our model; Section III presents a detailed analysis of comparative experiments alongside ablation studies for each module; Section IV showcases empirical tests and discusses our model's performance; and Section V provides a conclusive synthesis of our findings.

II. METHODS

The architectural framework of the DGSP-YOLO model, illustrated in Figure 1, closely resembles the structural design of YOLOv8n. The model is primarily composed of three key components: a backbone feature extraction network, a neck network, and a head detection module.

1) Backbone Network: The backbone of the DGSP-YOLO model has been carefully refined from the original YOLOv8 framework through the integration of advanced convolutional techniques, including Ghost convolution, the SPD-Conv module, and the C2fMHSA module. These enhancements significantly improve the network's ability for feature extraction and multi-scale semantic perception without substantially increasing the parameter count, thereby boosting detection precision for both large and small-scale ship targets in SAR imagery. Notably, the SPD-Conv module effectively addresses common challenges in SAR ship detection, such as detecting small targets and accommodating the varying resolutions of SAR imagery, by employing a spatial-to-depth transformation layer followed by a non-stepping convolutional layer. The C2fMHSA module, introduced before the feature fusion layer (SPPF), replaces the conventional C2f, allowing the model to focus more intensely on target features, which subsequently enhances detection accuracy.
2) Neck Section: The neck component of the DGSP-YOLO network preserves the structural essence of its YOLOv8 counterpart while introducing a pivotal innovation: the traditional upsampling layer has been replaced by the DySample module. This strategic substitution reduces the parameter count typically associated with the conventional upsampling layer (UpSample), all while maintaining an expedited inference rate and ensuring performance is not compromised.

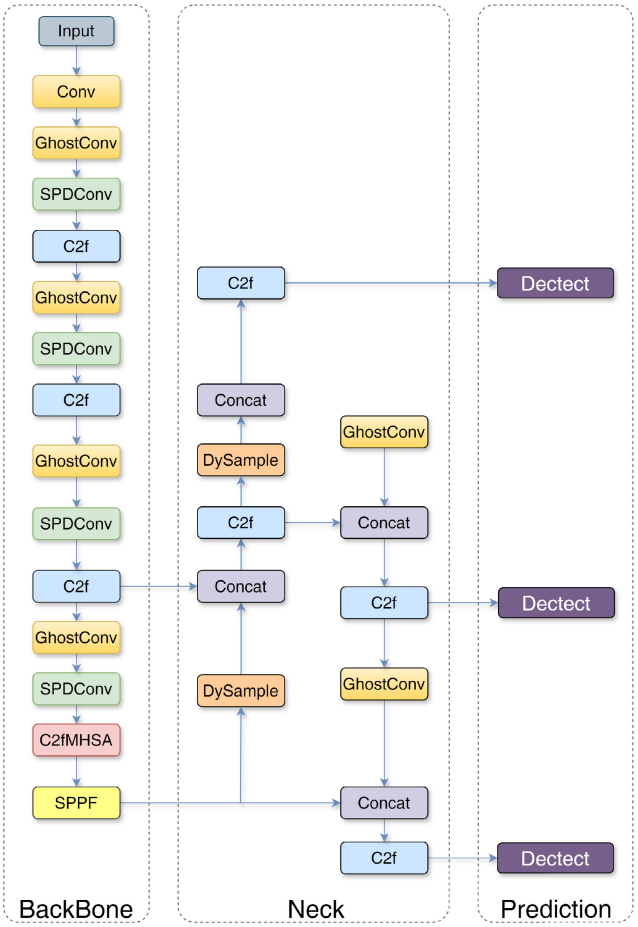


FIGURE 1. Overall architecture of the proposed DGSP-YOLO.

3) Head Section: The detection head of the DGSP-YOLO retains the original structural design of YOLOv8, featuring a decoupled architecture that separates the classification and regression heads. This modular approach promotes more effective feature learning and utilizes a feature aggregation strategy to integrate multi-scale features, thereby improving both detection precision and overall model effectiveness.

The subsequent sections of this paper are meticulously organized to delve deeper into the intricacies of our model's design, the comparative experimental analyses, and the implications of our findings on the field of SAR ship target detection.

A. BACKBONE IMPROVEMENTS

1) INTRODUCTION OF SPDConv MODULE

SAR imagery often encounters challenges such as low resolution, small target sizes, and complex background noise, with these issues becoming particularly pronounced during the detection of small-sized ship targets. These characteristics pose significant challenges to traditional target detection methodologies. In response to these challenges, this

study introduces a novel convolutional approach known as spatial depth transformed convolution (SPDConv) [36]. This method consists of a spatial-to-depth (SPD) layer followed by a non-band-step convolutional layer, as illustrated in Figure 2. The integration of SPDConv results in a significant enhancement in the detection of ship targets within SAR imagery, representing a notable advancement in the field of SAR-based ship detection.

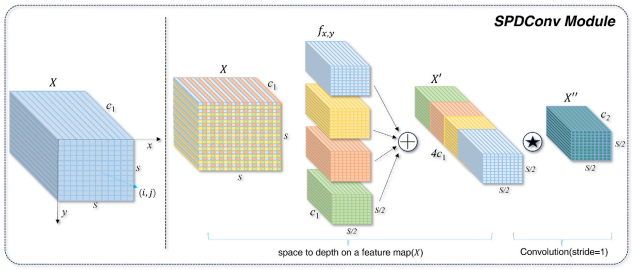


FIGURE 2. Details of SPDConv module.

SPDConv achieves efficient feature compression and information retention by effectively downsampling the feature maps within a convolutional neural network (CNN). The implementation details are as follows: first, given an intermediate feature map X with size $S \times S \times C_i$, it is downsampled using a specific slicing operation to generate a series of sub-feature maps $f_{x,y}$, which are then further processed to create the final feature representation.

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

The subgraph $f_{x,y}$ is meticulously constructed from elements $X(i, j)$ within the feature map that meet the criteria $i + x$ and $j + y$ being divisible by a predefined scaling factor, denoted as scale. In this context, i and j represent integer multiples of scale. To illustrate, when scale $= 2$, the feature map X is segmented into four distinct sub-feature maps. Each sub-map condenses its spatial extent by integrating along the channel axis, concomitantly expanding the channel dimension. This aggregation results in a novel feature map X', whose spatial dimensions are reduced to $\frac{1}{\text{scale}}$ of the initial size, with the channel dimensions expanding by a factor of scale^2.

Subsequently, after the spatial depth transformation, a non-stepping convolutional layer with a stride of 1 is employed, along with a C_2 convolutional filter. Here, C_2 represents a specified channel count that satisfies $C_2 < \text{scale}^2 < C_1$, facilitating the further conversion of X' to X''.

$$X'\left(\frac{S}{scale}, \frac{S}{scale}, scale^2 C_1\right) \rightarrow X''\left(\frac{S}{scale}, \frac{S}{scale}, C_2\right) \quad (2)$$

The primary rationale for employing non-spanning convolution in SAR ship detection is to preserve as much information as possible regarding all discernible features. Using a 3×3 filter with a stride of 3 may reduce the dimensions of the feature map; however, it means that each pixel is sampled only once, which could compromise information integrity. In contrast, employing a stride of 2 risks non-uniform sampling. In the context of SAR imagery, this can be particularly detrimental, as it may lead to blurring of the extracted features and a loss of critical details. For the precise detection of small targets, such as ships, these subtle details are crucial, directly influencing the accuracy and reliability of the detection process.

2) INTEGRATION OF THE C2fMHSA MODULE

Within the YOLOv8 framework, the C2f module plays a crucial role in enhancing the ability to capture intricate details and contextual information, thereby enriching the expressiveness of the extracted features. This improvement is achieved through the strategic combination of dual convolutional layers and advanced feature fusion techniques. Additionally, the incorporation of cross-layer connectivity within the module helps reduce redundant computational processes,significantly improving overall computational efficiency. The intricate design elements of the C2f module are illustrated in Figure 3.

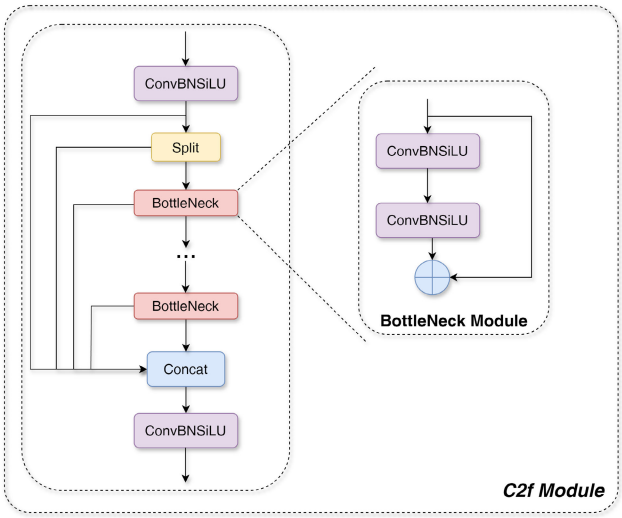


FIGURE 3. Details of C2f module.

While the C2f module has shown improvements in feature fusion, its traditional approach may not fully leverage the diverse levels of feature information in certain scenarios, thereby limiting the model's detection and generalization capabilities. To address this limitation, we introduce the C2fMHSA module, which incorporates an attention mechanism specifically designed to enhance the model's feature extraction capabilities. This innovative integration allows for a more nuanced analysis of multi-tiered feature data, thereby improving overall performance in SAR ship target detection tasks.

The attention mechanism enables the model to focus on the salient features of the target, thereby enhancing detection accuracy. The Multi-Head Self-Attention

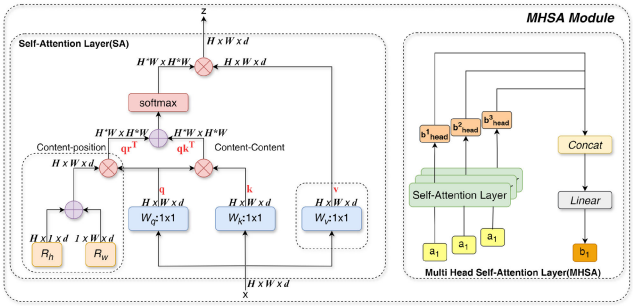


FIGURE 4. Details of MHSA attention.

mechanism (MHSA) [37] introduced in this study enriches feature representation through the synergistic operation of multiple attention heads, augmenting the model's expressiveness and robustness. The architecture of MHSA is illustrated in Figure 4. Within the operational framework of MHSA, input features are partitioned into h distinct subspaces, each conducting attention computations independently. The outputs from all attention heads are then consolidated and subjected to linear transformation to produce the final output. MHSA excels at capturing long-range dependencies within images, a capability crucial for ship detection due to challenges posed by intricate backgrounds, size variability, and partial occlusions that hinder conventional convolutional neural networks from achieving precise recognition. By integrating MHSA, the network can extract multi-scale ship feature information more efficiently, significantly enhancing detection accuracy. The detailed derivation of the MHSA module is presented subsequently:

1) First, given an input feature map, the input features are linearly transformed to obtain the query (Q), key (K) and value (V) matrices:

$$Q_i = QW_i^Q, K_i = KW_i^K, V_i = VW_i^V \tag{3}$$

where W_i^Q, W_i^K, W_i^V are trainable weight matrices.

2) Calculate the attention score:

$$Attention\,(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

3) Divide the input features into h subspaces (i.e., multiple heads) and compute attention independently for each subspace:

$$head_i = Attention\,(Q_i, K_i, V_i) \tag{5}$$

4) The outputs of all the attention heads are spliced together and linearly transformed to obtain the final output:

$$\begin{aligned} &MultiHead\,(Q, K, V) \\ &= Concat\,(head_1, head_2, \ldots, head_h)\,W^O \end{aligned} \tag{6}$$

where W^O is a trainable linear transformation matrix of the shape (d_{model}, d_{model}).

The integration of the multi-attention mechanism with the C2f module of the YOLOv8 architecture results in significant enhancements in SAR ship target detection. This hybrid module synergistically interacts with the Spatial Pyramid Pooling Feature (SPPF) module, facilitating a more efficient fusion of multiscale features. Such strategic integration enables the network to simultaneously discern ship targets across a range of sizes, with particular emphasis on improving the detection accuracy of smaller vessels. This approach is crucial for advancing the performance metrics of SAR-based ship detection systems.

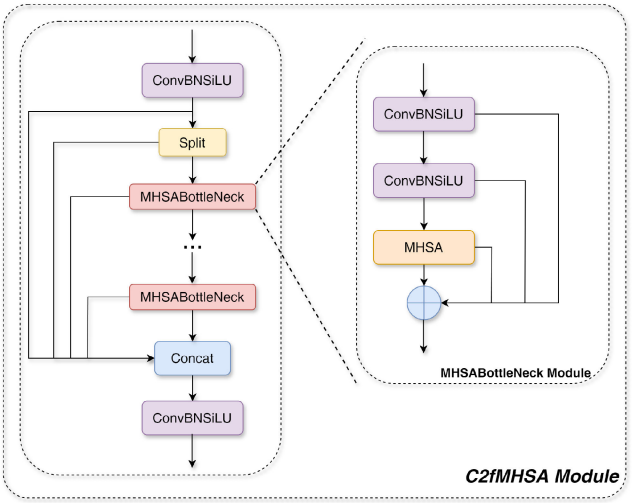


FIGURE 5. Details of C2fMHSA module.

The C2fMHSA module initially processes the input feature map by integrating local feature extraction through the C2f structure with global context modeling via the MHSA mechanism. The output of this module is a feature map enriched with both local and global dependencies. This feature map is then fed into the SPPF module, which applies spatial pyramid pooling at various scales (e.g., 1×1, 3×3, 5×5) to capture multi-scale spatial information. The outputs from these pooling operations are concatenated to form a multi-scale feature representation, which is subsequently utilized for further processing within the network. This interaction enables the network to leverage both global attention and multi-scale spatial features, enhancing its capability to manage objects of varying sizes and complex spatial relationships.

Furthermore, the integration of the MHSA mechanism into the C2f module of the YOLOv8 framework effectively mitigates the influence of extraneous background elements. This is accomplished by assigning varying levels of importance to distinct features, thereby guiding the network's focus towards the critical targets of interest. The MHSA mechanism significantly enhances the network's feature extraction capacity and improves the model's resilience in complex background environments. The schematic design and operational principle of the C2fMHSA module are illustrated in Figure 5. This module provides robust support for the detection of ships within SAR imagery through a

meticulously crafted feature processing workflow, ensuring a higher level of detection accuracy and reliability.

3) IMPLEMENTATION OF THE GhostConv MODULE

In the domain of target detection, convolutional layers (Conv) play a crucial role in extracting essential image features such as edges, textures, and colors, which are vital for accurate target identification. However, the traditional convolutional approach in YOLOv8 can lead to increased computational costs, as the parameter count scales with the dimensions of the output feature map. To address this challenge, we introduce GhostConv [38], an innovative optimization of the convolutional layer. This method enhances the network's representational capacity by generating additional "ghost" feature maps while concurrently reducing the parameter count. The feature extraction process begins with a conventional convolution, followed by an efficient linear transformation to create a derived feature map. This dual-step approach improves the efficiency and accuracy of feature extraction while minimizing computational overhead. This pioneering technique enables YOLOv8 to significantly enhance the model's performance and efficiency without compromising recognition accuracy.

The detailed implementation of GhostConv is as follows: Initially, a 1×1 convolutional filter is deployed to condense the channel count of the input feature map, thereby generating an intrinsic feature map. This step minimizes redundant computation attributed to the channel reduction. The subsequent equation elucidates this process.

$$Y' = X * f' \tag{7}$$

Let $X \in \mathbb{R}^{c\times h\times w}$ denote the input feature map, where $f \in \mathbb{R}^{c\times k\times k\times m}$ represents the convolution filter. The output $Y' \in \mathbb{R}^{h'\times w'\times m}$, yielded by the primary convolution, comprises a set of m feature maps. Here, h and w signify the height and width, respectively; c indicates the total input channel count; m represents the total output channel count; and k is the kernel size of the convolution. Subsequently, to procure the desired n feature maps, each intrinsic feature map of Y' undergoes a low-cost linear transformation, resulting in s ghost feature maps, as delineated in the following equation:

$$y_{ij} = \Phi_{i,j}\left(y'_i\right), \forall i = 1, \ldots, m, j = 1, \ldots, s \tag{8}$$

where $\Phi_{i,j}$ is denoted as a linear transformation and y_{ij} is meant to be the generated ghost feature map. This approach increases the feature set more efficiently while minimizing redundancy. The GhostConv module is depicted in the Figure 6.

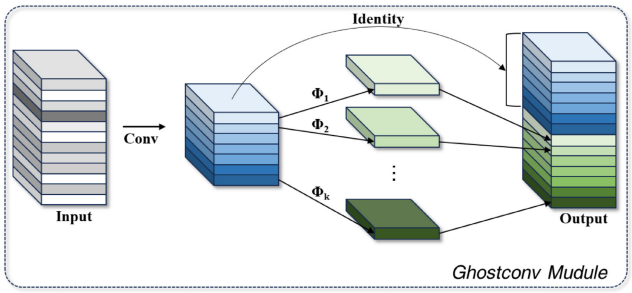


FIGURE 6. Details of GhostConv module.

B. IMPROVEMENTS IN THE NECK REGION

1) DySample MODULE

In the canonical YOLOv8 architecture, the up-sampling layer, located within the neck region, is responsible for increasing the resolution of downsampled, low-resolution feature maps, thereby aligning them with the higher-resolution outputs produced by the Backbone. This up-sampling process is essential for recovering spatial details and retaining high-level semantic information. However, conventional up-sampling techniques often fall short, as they can introduce noise or blurring that compromises the integrity of feature maps, while also increasing computational demands and slowing down model inference.

To address these limitations, this paper introduces DySample [39], an innovative dynamic upsampling module. DySample transforms the upsampling paradigm by employing a point sampling strategy, thereby circumventing the computational overhead associated with kernel-dependent upsampling methods. Unlike existing dynamic upsampling modules such as CARAFE [40], FADE [41], and SAPA [42], DySample eliminates the need for high-resolution guiding feature inputs and does not require additional CUDA support. This innovation strikes an optimal balance between performance and efficiency, significantly reducing the experimental burden in scientific research endeavors. The architectural design and operational mechanism of the DySample module are illustrated in Figure 7.

The procedural design elements of the DySample module are concisely outlined as follows:

The generation of offsets commences with a linear layer, which is followed by the incorporation of a dynamic range factor $\sigma \in [0, 0.5]$ to enhance the adaptability of the offsets. This approach is delineated in the subsequent equation:

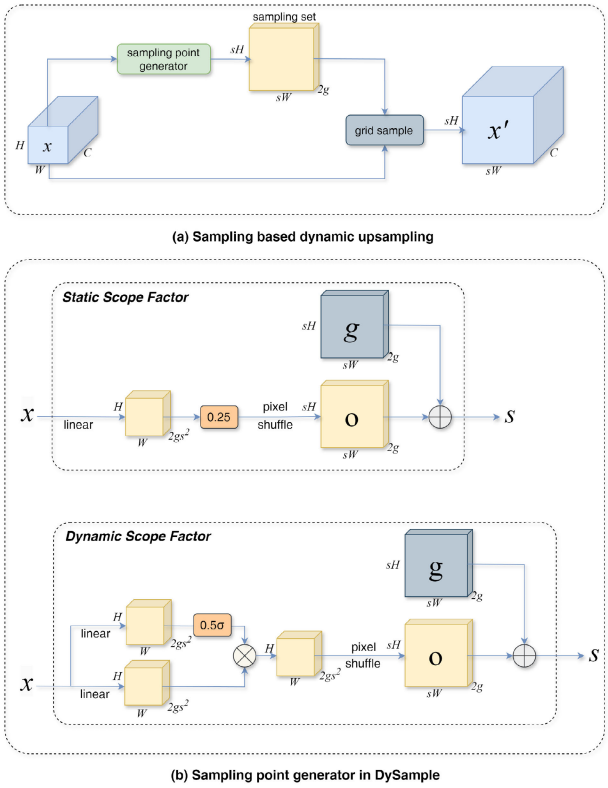
$$O = 0.5\text{sigmoid}\left(\text{linear}_1\left(X\right)\right) \cdot \text{linear}_2\left(X\right) \tag{9}$$

The aforementioned offsets are subsequently deployed to resample the progressively interpolated feature maps. This resampled output represents a synthesis of the offsets O and the initial sampled network G. Thereafter, this sampled output is repurposed to further resample the feature maps, as articulated in the following equation:

$$S = O + G \tag{10}$$

Ultimately, the resultant feature map X' is acquired through the process of bilinear interpolation, as articulated in the accompanying equation:

$$X' = \text{grid\_sample}\left(X, S\right) \tag{11}$$



(a) Sampling based dynamic upsampling

(b) Sampling point generator in DySample

FIGURE 7. Details of DySample module.

C. LOSS FUNCTION

The BBR loss function is crucial in the field of ship detection. In this study, we introduce a novel loss function called EIoU [43], specifically designed for the recognition of synthetic aperture radar (SAR) ship images. The EIoU loss function enhances the conventional CIoU loss function by incorporating a broader set of geometric parameters, thereby improving the optimization process in terms of both efficiency and stability. This refinement leads to reduced training errors and accelerated convergence rates. Notably, the EIoU loss function significantly mitigates the effects of insufficient gradient information, particularly in scenarios where bounding box overlap is minimal. This characteristic enhances the model's generalization capabilities, enabling it to accurately identify ship targets across various background contexts.

The BBR loss function, which is computed based on the Intersection over Union (IoU), is defined as follows:

$$L_{IoU} = 1 - IoU = 1 - \frac{W_i H_i}{S_i} \tag{12}$$

$$W_i = \max\left(0, \min\left(x2_{pre}, x2_{gt}\right) - \max\left(x1_{pre}, x1_{gt}\right)\right) \tag{13}$$

$$H_i = \max\left(0, \min\left(y2_{pre}, y2_{gt}\right) - \max\left(y1_{pre}, y1_{gt}\right)\right) \tag{14}$$

Within this framework, as depicted in Figure 8, S_i denotes the intersecting area of the prediction box with the real box, with H_i and W_i signifying the height and width of the overlapping region, respectively.

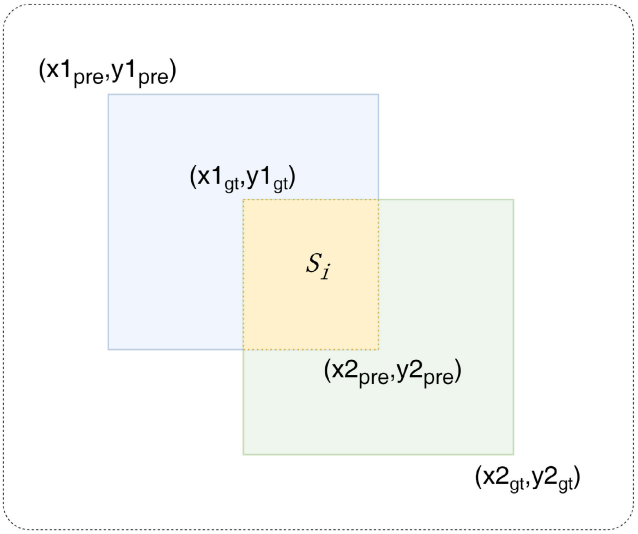


FIGURE 8. Overlap area diagram. The yellow area represents the overlapping part.

The scenario where $H_i = 0$ or $W_i = 0$ typically suggests that the prediction box has degenerated into a point or a line segment, culminating in an overlap area of zero with the actual bounding box. Under such circumstances, the conventional BBR loss L_i is prone to the challenge of vanishing gradients.

To address the gradient vanishing issue, enhancements have been progressively integrated into the Intersection over Union (IoU) loss function. The Generalized IoU (GIoU) [44]introduces a term that accounts for the smallest enclosing box, while the Distance IoU (DIoU) [45]incorporates a centroid Euclidean distance measure. Furthermore, the Complete IoU (CIoU) includes an aspect ratio discrepancy, and the Shape IoU (SIoU) [46]focuses on shape dissimilarity, thereby enhancing the precision of shape alignment.

An effective loss function must impose appropriate geometric penalties when the predicted bounding box intersects with the actual bounding box. Throughout the training phase, it is crucial to carefully consider the potential adverse impacts of low-quality samples within the dataset, as this can bolster the model's generalization capabilities and minimize the need for interventions. Such samples, influenced by factors such as the aspect ratio of the annotated frames and their proximity to the target, may hinder the model's generalization. In the context of SAR imagery, ship target recognition poses challenges due to the diverse shapes and sizes of ships, as well as the complexity of the background. To address these challenges, the Enhanced IoU (EIoU) introduces three geometric penalty terms to improve the loss function's effectiveness. The EIoU formulation can be articulated as follows:

$$L_{EIoU} = L_{IoU} + L_{\text{dis}} + L_{\text{asp}} \tag{15}$$

$$L_{dis} = \frac{\rho^2\left(b, b^{gt}\right)}{(w^c)^2 + (h^c)^2} \tag{16}$$

$$L_{asp} = \frac{\rho^2\left(w, w^{gt}\right)}{(w^c)^2} + \frac{\rho^2\left(h, h^{gt}\right)}{(h^c)^2} \tag{17}$$

$$L_{EIoU} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{(w^c)^2 + (h^c)^2} + \frac{\rho^2\left(w, w^{gt}\right)}{(w^c)^2} + \frac{\rho^2\left(h, h^{gt}\right)}{(h^c)^2} \tag{18}$$

where IoU denotes the intersection and concurrency ratio between the predicted and actual boxes, ρ symbolizes the Euclidean distance, and b and b^{gt} represent the centroids of the predicted and actual bounding boxes, respectively. The dimensions w, h, w^{gt}, and h^{gt} correspond to the widths and heights of the predicted and actual boxes, respectively, while w^c and h^c denote the widths and heights of the smallest enclosing bounding rectangles that encompass both boxes.

Furthermore, L_{dis}, termed the distance loss or centroid distance penalty, is calculated by taking the square of the Euclidean distance between the centroids of the predicted box b and the true box b^{gt}, $\rho^2(b, b^{gt})$, and normalizing it by the square of the diagonal of the smallest enclosing rectangle, $(w^c)^2 + (h^c)^2$. This component is instrumental in providing gradient information in instances where the bounding boxes lack overlap, thereby guiding the model to refine the positioning of the predicted box.

L_{asp}, referred to as the aspect ratio loss, includes two penalty terms that assess the discrepancies in width and height between the predicted box and the true box. Each term is computed as the square of the difference between the predicted and actual dimensions, ρ^2, normalized by the square of the width $(w^c)^2$ and the square of the height $(h^c)^2$ of the smallest enclosing rectangle, respectively. This aspect ratio loss incentivizes the model to accurately learn the width-to-height ratio, ensuring effective gradient provision even when bounding boxes do not intersect.

By integrating the traditional IoU loss with these additional geometric penalty terms—distance loss and aspect ratio loss—the EIoU loss function is designed to enhance the convergence rate and accuracy of the model in bounding box regression tasks. This is particularly beneficial in scenarios involving small targets or bounding boxes with minimal overlap, thus optimizing the detection performance in SAR ship target detection.

III. RESULTS

To validate the efficacy of the DGSP-YOLO model in the domain of synthetic aperture radar (SAR) ship target detection, we conducted comprehensive experimental validation across three significant public datasets: HRSID [47], LS-SSDD-v1.0 [48], and SSDD [49]. This section of the paper first outlines the experimental setup, the datasets used, the evaluation metrics, and the details of the training process. Following this, ablation studies are performed to ascertain the contribution of each component module. Subsequently, a comparative analysis positions the proposed DGSP-YOLO model against the current state-of-the-art (SOTA) detectors in the field.

A. EXPERIMENTAL SETUP

In our experiments, we use YOLOv8 as the baseline model, the dataset is in TXT format of the YOLO model, the batch size is set to 30, the number of training rounds is 300 epoch, the SGD optimizer is used, and the initial learning rate is used to be 0.01 and the momentum is 0.927. All the experiments are carried out on a 4060ti GPU. The software and hardware environments required for the experiments are shown in Table 1.

TABLE 1. Software and hardware requirements for the experiment.

Component	Specifications
CPU	Intel(R) Core(TM) i5-12600KF
RAM	32GB
GPU	NVIDIA GeForce RTX 4060Ti
Language	Python 3.8
Framework	PyTorch 2.2.2
Operating System	Ubuntu 22.04

B. DATASETS

In this study, we use three datasets, HRSID, LS-SSDD-v1.0 and SSDD, to evaluate the performance of the DGSP-YOLO model, where the dataset division ratio is 7:2:1 for the training set: validation set: test set. The specific parameters of the three datasets are shown in Table 2. The images sizes is the size set in the experiment.

1) HRSID DATASET

The HRSID dataset was introduced by Hao Su of the University of Electronic Science and Technology (UEST) in January 2020 as a comprehensive resource for ship detection, semantic segmentation and instance segmentation in SAR images. This dataset comprises 5,604 images, encompassing 16,965 annotated ship instances. The images are formatted at a size of 800 × 800 pixels. The ship target annotations in HRSID are depicted in Fig. 9a, and representative sample images are shown in the subsequent Figure 10.

2) LS-SSDD-v1.0 DATASET

Assembled by Professor Xiaoling Zhang's team at the University of Electronic Science and Technology (UEST), the LS-SSDD-v1.0 dataset is tailored for the detection of small vessels within large-scale SAR images. This dataset comprises 9,000 images, extracted from 15 extensive SAR images, totaling 6,015 ship targets. Each image in the dataset is formatted at a size of 800 × 800 pixels. The annotations for ship targets in LS-SSDD-v1.0 are illustrated in Figure 9b, with exemplary images provided in the accompanying Figure 11.

TABLE 2. Specific parameters of the three datasets.

Datasets	HRSID	LS-SSDD-v1.0	SSDD
numbers	5604	9000	1160
Ship numbers	16965	6015	2551
Images sizes	800×800	800×800	500×500

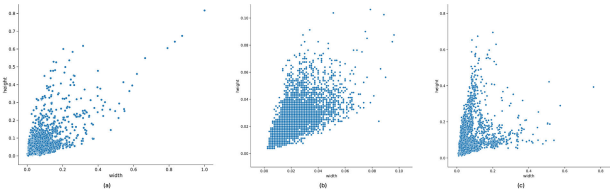


FIGURE 9. Ship scale information on three datasets: (a) distribution of ship targets on HRSID; (b) distribution of ship targets on LS-SSDD; (c) distribution of ship targets on SSDD.

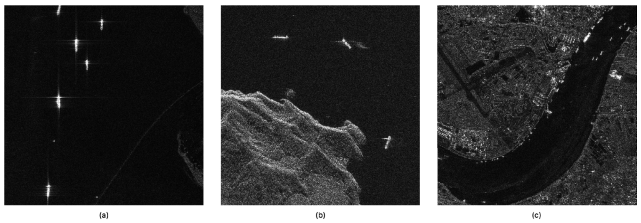


FIGURE 10. Samples of HRSID dataset. (a) Multiscale ship samples, (b) inshore ship samples, and (c) small ship samples.

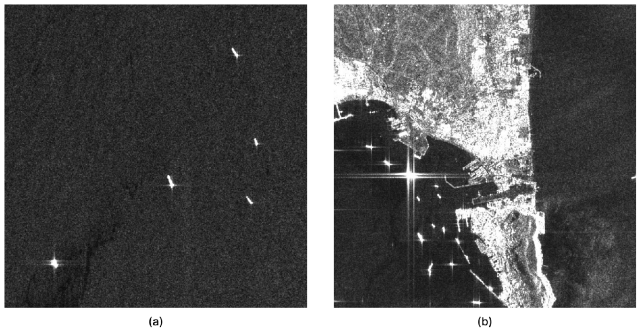


FIGURE 11. Samples of LS-SSDD-v1.0 dataset. (a) Offshore ship samples, and (b) inshore ship samples.

3) SSDD DATASET

Crafted by the Department of Electronics and Information Engineering at Naval Aeronautics and Astronautics University (NAAU), the SSDD dataset is specifically designed for ship detection tasks. It encompasses 1,160 images with 2,456 ship targets. The sizes of the original images in the dataset were not uniform, but in this experiment we resized all images to a uniform size of 500 × 500 pixels to ensure consistency and ease of processing. The distribution of aspect ratio information for ship targets within the SSDD dataset is detailed in Figure 9c, with a selection of images from the dataset displayed in the Figure 12.

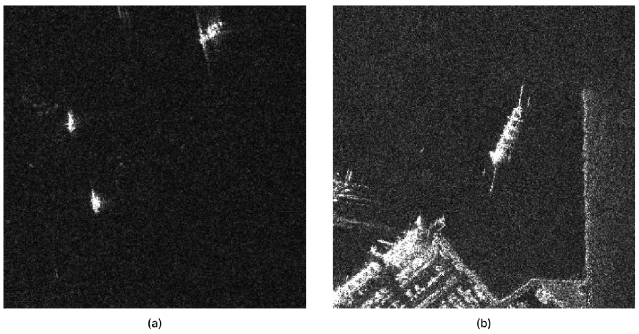


FIGURE 12. Samples of SSDD dataset. (a) Offshore ship samples, and (b) inshore ship samples.

C. EVALUATION METRICS

To conduct a comprehensive evaluation of the DGSP-YOLO model's performance in SAR ship target detection, we utilize three key metrics: precision (P), recall (R), and mean average precision (mAP). Precision (P) quantifies the proportion of detections accurately identified as ships relative to all positive identifications, while recall (R) measures the proportion of actual ships correctly identified compared to the ground truth data. These metrics are derived from the following components: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). TP represents the correctly identified ship targets, FP denotes the incorrectly identified non-ship targets classified as ships, TN signifies the correctly identified background areas as non-targets, and FN refers to the missed ship targets that were incorrectly classified as background. Consequently, precision (P) is calculated as the ratio of TP to the sum of TP and FP, expressed as:

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

The recall rate R is the number of all vessels detected as a percentage of the total number of vessels and is expressed as:

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

AP is the area bounded by the PR curve, expressed as:

$$AP = \int_0^1 P(R)\, dR \tag{21}$$

The mAP is the mean of the AP values of multiple categories, and since there is only one ship category in this study, where N = 1, the mAP value is equal to the AP, which is expressed as:

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i = \int_0^1 P(R)\, dR \tag{22}$$

D. ABLATION EXPERIMENT

To ascertain the individual contributions of each module to the model's detection capabilities, we conducted ablation

studies. Initially, we selected the publicly available HRSID dataset as a benchmark for evaluating the impact of each module on detection performance compared to the YOLOv8 model. The performance metrics employed for this assessment include precision (P), recall (R), mAP50, and mean average precision across IoU thresholds from 50% to 95% (mAP50:95). The configuration of the ablation experiments is presented in Table 3, where the modules are denoted as follows: A for SPDConv, B for DySample, C for C2fMHSA, D for GhostConv, and E for EIoU.

The analysis presented in the table reveals that the incorporation of the SPDConv module significantly bolsters the baseline model's performance across all evaluated metrics, with particularly notable enhancements in precision (P) and mAP50:95, which exhibit improvements of 1% and 0.7%, respectively. These gains are attributed to SPDConv's superior detection capabilities for small targets. The integration of the dynamic up-sampling module, DySample, allows the model to preserve richer semantic information during the up-sampling process, thereby enhancing the model's recognition accuracy. The inclusion of the C2fMHSA module augments the model's feature extraction capabilities. While there is a slight decrease in the accuracy rate, this module leads to improvements in the recall rate, mAP50, and mAP50:95. The addition of the GhostConv module enables the model to obtain a greater number of feature maps at a reduced computational cost, particularly beneficial for scenarios involving small targets, and results in a 2.1% increase in mAP50:95 compared to the preceding model iteration.Furthermore, the adoption of the EIoU loss function has yielded a marked improvement in mAP, achieving a rate of 94%. This enhancement is due to EIoU's ability to mitigate the detrimental effects of low-quality samples on detection performance by offering a more comprehensive similarity metric, which allows for a more precise assessment of the alignment between predicted and target bounding boxes.

E. COMPARATIVE EXPERIMENTS

1) VALIDATION OF THE C2fMHSA

Within the backbone of the YOLOv8 benchmark model, we integrated the Multi-Head Self-Attention (MHSA) module alongside the Channel and Spatial Feature Fusion (C2f) module, preceding the SPPF module. This integration enhances the backbone's feature extraction capabilities, synergistically improving the network's ability to capture the subtleties of SAR imagery through SPPF. To substantiate the efficacy of the C2fMHSA module, a comparative analysis was conducted with other prevalent attention modules. Specifically, we assessed the performance of the initial module against traditional attention mechanisms, including Convolutional Attention (CA) [50], Squeeze-and-Excitation (SE) [51], Efficient Channel Attention (ECA) [52], and Convolutional Block Attention Module (CBAM) [53], across three datasets: HRSID, LS-SSDD-v1.0, and SSDD. The comparative evaluation, incorporating diverse attention mechanisms into the benchmark model, is presented in Table 4.

The tabulated results clearly demonstrate the improved performance of the model integrated with the MHSA attention module. Notably, on the HRSID ship dataset, there is a marginal yet significant increase in accuracy (P) by 0.4% and in mAP50 by 0.2% compared to the baseline model; however, no substantial change was observed in recall. In the case of the LS-SSDD-v1.0 dataset, while recall (R) experienced a slight decline of 1.5%, accuracy (P) saw a notable rise of 2.2%, and mAP50 increased by 1.3%. On the SSDD dataset, the model achieved a 1.7% increase in accuracy, a 4.6% increase in recall, and a 0.4% increase in mAP50. The incorporation of the MHSA attention module allows the model to focus more intently on critical regions, thereby enhancing its overall performance. Comparative analysis with other attention mechanisms reveals that MHSA not only performs admirably but also excels due to its robust feature extraction capabilities and adaptability in various ship recognition scenarios.

2) COMPARISON OF DIFFERENT LOSS FUNCTIONS

To assess the efficacy of the EIoU loss function in the context of SAR ship target detection, we conducted a comparative analysis against three variants of the Basic Box Regression (BBR) loss functions: GIoU, CIoU, DIoU, and SIoU. This evaluation was performed using the HRSID dataset, with the experimental outcomes presented in Table 5.

The data presented in the table clearly delineates the following outcomes: On the HRSID dataset, the model utilizing the CIoU loss function achieved the highest accuracy, reaching 94.1%. However, its recall and mAP values were surpassed by those of the EIoU loss function. The EIoU loss function not only maintained a higher accuracy but also attained a recall rate of 86.9%, which is an improvement over the DIoU loss function's recall rate of 87.2%. Although the DIoU loss function demonstrated the best recall rate, its accuracy and mAP50 stood at 92.2% and 93%, respectively, still falling short compared to EIoU. Collectively, EIoU's mAP50 value reached 94%, outperforming all other loss functions in comparative experiments on the HRSID dataset.

It is particularly noteworthy that on the LS-SSDD-v1.0 dataset, which comprises small-scale ship targets, the EIoU loss function surpassed other loss functions in terms of recall and mAP50, with rates of 62.1% and 69%, respectively. This indicates that the incorporation of the EIoU loss function significantly reduces the model's false negative rate when detecting small-scale ship targets.

Furthermore, experimental results on the SSDD dataset demonstrated accuracy (P) and mAP50 values of 97.4% and 99%, respectively. These outcomes substantiate the model's robust generalization capabilities and comprehensive performance when equipped with the EIoU loss function for SAR ship detection tasks.

TABLE 3. Results of ablation experiments.

Model	A	B	C	D	E	P	R	mAP50	mAP50:95
YOLOv8	-	-	-	-	-	92.2	86.4	93	69.8
	√	-	-	-	-	93.2	86.4	93.3	70.5
	√	√	-	-	-	94.3	85.8	93.1	70.1
	√	√	√	-	-	92.6	87	93.6	70.1
	√	√	√	√	-	94.1	86.8	93.5	72.2
	√	√	√	√	√	93	86.9	94	72.2

TABLE 4. Performance comparison of different methods.

Model	Methods	HRSID			LS-SSDD-v1.0			SSDD		
		P	R	mAP50	P	R	mAP50	P	R	mAP50
YOLOv8	C2f	92.2	86.4	93	78.9	61.9	66	95.2	95.3	98.5
	C2fCA	90.9	85.2	92.3	80	61	66.5	97.6	94.9	98.4
	C2fSE	91.1	87.3	92.8	81.7	57.7	65.5	95.3	96.4	98.2
	C2fECA	92	85.3	92.3	76.1	61.4	66.4	97.4	95.8	98.9
	C2fCBAM	89.7	86.7	92.5	80.9	59.5	66.5	97.4	94.6	98.3
	C2fMHSA	**92.6**	86.1	**93.2**	81.1	60.4	**67.3**	96.9	**97.9**	**98.9**

TABLE 5. Performance comparison of different loss functions.

Model	IoU	HRSID			LS-SSDD-v1.0			SSDD		
		P	R	mAP50	P	R	mAP50	P	R	mAP50
DGSP-YOLO	GIoU	93.1	86.9	93.1	79.1	61.8	66.3	96.6	97.5	98.8
	DIoU	92.2	87.2	93	81.6	59.5	65	96.9	95.9	98.8
	SIoU	92.2	86.9	92.9	85.9	60.3	67	96.1	98.1	98.7
	CIoU	94.1	86.8	93.5	83.6	61.6	66.4	97.1	94.4	98.5
	EIoU	93	**86.9**	**94**	83.7	**62.1**	**69**	**97.4**	96.4	**99**

3) COMPARISON OF DIFFERENT DETECTORS

For a more comprehensive assessment of the DGSP-YOLO model's performance, four key metrics were employed to evaluate its efficacy across three datasets: HRSID, LS-SSDD-v1.0, and SSDD. These metrics include precision (P), recall (R), mAP50, and mAP50:95. The input images from the three datasets were uniformly resized to 800 × 800 pixels to ensure consistency throughout the experiments. Comparative experiments were also conducted with several mainstream detectors, including the two-stage detector Faster R-CNN, the single-stage convolutional SSD, various versions of the YOLO model, and the Transformer-based RT-DETR [54] detector. Table 6 presents a comparative analysis of DGSP-YOLO's experimental outcomes against those of other prominent target detection algorithms.

The data presented in the tables unequivocally demonstrate the enhanced detection capabilities of DGSP-YOLO when evaluated across the HRSID, LS-SSDD-v1.0, and SSDD datasets. Relative to the baseline model, DGSP-YOLO has achieved significant improvements in both mAP50 and mAP50:95 metrics—on HRSID, increasing from 93% and 69.8% to 94% and 72.2%; on LS-SSDD-v1.0, improving from 66% and 24.2% to 69% and 25.3%; and on SSDD, rising from 98.5% and 71.8% to 99% and 75.1%. On the HRSID dataset, DGSP-YOLO notably attains a recall rate of 86.9%, surpassing other detectors and underscoring its relevance in ship detection scenarios. Moreover, it excels in mAP50 and mAP50:95, reaching 94% and 72.2%, respectively, indicating a substantial improvement over other algorithms, particularly in detecting ships at various scales.

On the LS-SSDD-v1.0 dataset, which features smaller ship targets and thus poses greater detection challenges, DGSP-YOLO maintains a commendable accuracy of 83.7% and a robust recall of 62.1%. It also achieves key metric scores of 69% for mAP50 and 25.3% for mAP50:95, surpassing other detection models. In the SSDD dataset evaluation, DGSP-YOLO achieves an accuracy of 97.4% and a recall of 96.4%, with mAP50 and mAP50:95 reaching 99% and 75.1%, respectively. These results are the most outstanding among all compared models and further confirm DGSP-YOLO's generalization and robustness across different datasets.

The PR Figure 13 presents a comparison of the performance of mainstream YOLO series detection models, where DGSP-YOLO occupies a larger area, indicating broader coverage in terms of precision and recall. Considering the experimental results across three datasets, the superior performance of DGSP-YOLO in ship detection tasks is evident. It outperforms existing mainstream detectors in precision, recall, and mean Average Precision (mAP). These impressive results can be attributed to the incorporation of SPDConv and attention modules, as well as improvements to the overall model architecture, which facilitate more effective extraction of ship targets and enhance the model's detection capabilities. Furthermore, the adoption of the EIoU loss

TABLE 6. Performance comparison of different detectors.

Model	HRSID				LS-SSDD-v1.0				SSDD			
	P	R	mAP50	mAP 50:95	P	R	mAP50	mAP 50:95	P	R	mAP50	mAP 50:95
Faster R-CNN	85.1	86.5	86.1	64	76.2	78.3	59.5	21.2	93.2	94.5	94.2	64.6
SSD	92.3	58.68	79.82	50.3	78.62	50.59	56.29	20.2	94.20	72.76	94.4	57
YOLOv5n	91.5	85.4	92.2	69	79.6	61.2	65.8	23.8	95.8	94.6	98.5	71.3
RT-DETR	84	73.4	83.2	54.3	70	51.6	55.8	18.8	92.4	85.1	95.5	67.1
YOLOv3-tiny	90.8	77.1	86.3	62.8	75.5	59.7	64.2	23.4	94.2	92.7	97.3	72.9
YOLOv7-tiny	86.7	71	81.1	54.7	72.8	57.7	60.8	21.6	92.7	88.9	94.2	62.5
YOLOv6	92.5	81.9	91.2	66.2	78.7	59.4	63.7	22.8	97.1	**97.1**	98.8	72.8
YOLOv9-tiny	93.3	85.6	92.8	69.6	81.8	60.4	65.9	23.8	95.1	96.5	98.7	74.7
YOLOv8n	92.2	86.4	93	69.8	78.9	61.9	66	24.2	95.2	95.3	98.5	71.8
YOLOv10n	**93.5**	85.4	92.6	71.1	79.3	61	66.3	23.9	97.1	95.1	98.4	72.7
DGSP-YOLO	93	**86.9**	**94**	**72.2**	**83.7**	**62.1**	**69**	**25.3**	**97.4**	96.4	**99**	**75.1**

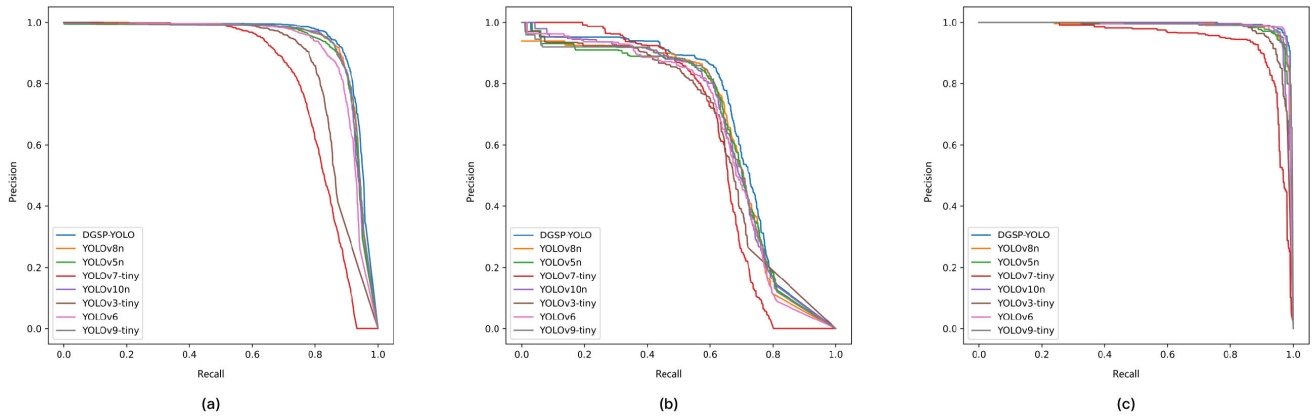


FIGURE 13. Precision-Recall Curve. (a) HRSID datasets, (b) LS-SSDD-v1.0 datasets, and (c) SSDD datasets.

function refines the model's object bounding box predictions, thereby enhancing its generalization and robustness.

IV. DISCUSSION

To rigorously assess the efficacy of our approach for detecting small-scale ship targets across various dimensions in real-world scenarios, we deliberately selected a subset of images for evaluation.

Specifically, we extracted three representative images from each of the three comprehensive datasets: HRSID, LS-SSDD-v1.0, and SSDD. Subsequently, we conducted inference validation using both the YOLOv8 and DGSP-YOLO models. The outcomes of these inference assessments are graphically depicted in Figure 14.

Figure 14(a) illustrates the ground truth annotations, while Figures 14(b) and 14(c) depict the detection inferences of YOLOv8 and DGSP-YOLO, respectively. In Figure 14(a1), where the image background is predominantly oceanic with minor trailing noise interference, both YOLOv8 and DGSP-YOLO exhibit commendable performance. However, the confidence levels of the ship targets identified by YOLOv8 are generally lower than those of DGSP-YOLO. In Figure 14(a2), which features an image with an entirely oceanic background and smaller-sized targets, YOLOv8 fails to detect one of the small ship targets, as indicated by the yellow circle. This observation confirms YOLOv8's lower recall rate on the LS-SSDD-v1.0 dataset. In contrast, DGSP-YOLO accurately detects all 11 ships present, demonstrating superior confidence levels compared to YOLOv8. Thus, it can be concluded that DGSP-YOLO maintains a significant advantage in detecting small targets. Figure 14(a3) presents the detection outcomes of both algorithms in a near-shore scenario, which is challenging due to the complex background and numerous interferences. YOLOv8 erroneously classifies a shore object as a ship, marked by the blue circle, whereas DGSP-YOLO demonstrates superior performance, not only in confidently detecting large-sized ship targets but also in accurately identifying small-sized ship targets. DGSP-YOLO effectively discriminates genuine ship targets from the noisy near-shore background, without any false positives or omissions. Furthermore, the heightened recognition confidence on the SSDD dataset is attributed to the presence of more large-sized targets and the dataset's enhanced adaptability relative to LS-SSDD-v1.0 and HRSID. In summary, DGSP-YOLO's performance is remarkable, surpassing the benchmark model YOLOv8 across all three datasets, thereby validating its efficacy and robust generalization capabilities in SAR ship detection.

To ascertain the nature of the features that our model has learned and to evaluate whether these features align with our expectations or if the model has inadvertently learned to exploit spurious cues, we employ a visualization

technique that renders the model's gradient computations within the image as a heatmap. Gradient-weighted Class Activation Mapping (Grad-CAM) [55] plays a crucial role in this analysis, as it isolates the model's focus on specific regions of interest, providing insights into the information that the network prioritizes.

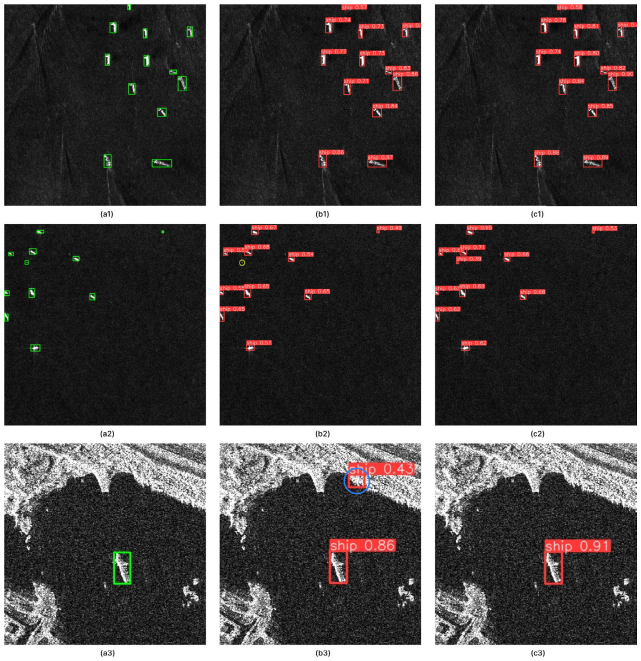


FIGURE 14. Visual detection results on HRSID, LS-SSDD-v1.0, and SSDD datasets. (a1), (a2) and (a3) are denoted as the ground truth of HRSID, LS-SSDD-v1.0, SSDD, respectively, (b1), (b2) and (b3) are denoted as the detection results of YOLOv8n, and (c1), (c2) and (c3) are denoted as the detection results of DGSP-YOLO.

This method allows us to assess whether the network is genuinely learning the intended features and information. We utilized Grad-CAM to visualize the gradient information extracted from the feature maps by the model from the images across the three datasets. The results of this visualization are depicted in Figure 15, which elucidates the model's decision-making process and the relevance of the learned features to the task of SAR ship detection.

In the heatmap visualization, the intensity of the red color block is indicative of the model's focus on specific image areas, with more intense red signifying a higher concentration of extracted feature information. In Figure 15(a1), which utilizes offshore samples from the HRSID dataset, it is evident that YOLOv8 is heavily concentrated on environmental features not directly associated with the ship, while the DGSP-YOLO model demonstrates a more targeted focus on the ship itself, with the red areas predominantly highlighting the ship's hull. Figure 15(a2), representing a nearshore image from the LS-SSDD-v1.0 dataset, reveals that due to the dataset's prevalence of small-scale targets, the heatmaps generated by YOLOv8 exhibit a greater number of non-red regions, suggesting potential challenges in detecting smaller targets. In contrast, DGSP-YOLO maintains strong performance, with the red color consistently covering the entire target area, indicating its superior capability in capturing target information for small-scale target detection and thus enhancing identification accuracy. Figure 15(a3), an image from the SSDD dataset showcasing a significant size variation in ship targets, demonstrates that our method, DGSP-YOLO, effectively focuses on the ship targets themselves, rather than irrelevant details, irrespective of whether the target is large or small. This consistent focus on relevant target areas underscores DGSP-YOLO's exceptional performance in the detection of ships within SAR imagery.

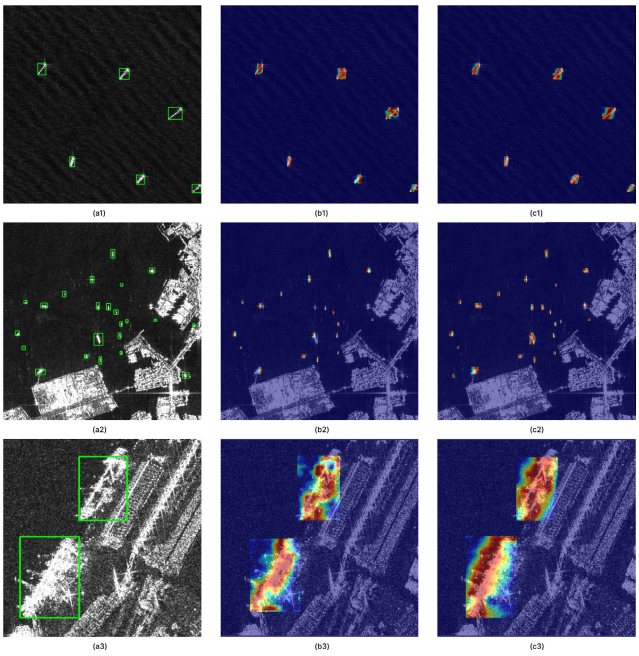


FIGURE 15. Grad-CAM visualization results of HRSID, LS-SSDD-v1.0, and SSDD datasets. (a1), (a2) and (a3) are denoted as the ground truth of HRSID, LS-SSDD-v1.0, SSDD, respectively, (b1), (b2) and (b3) are denoted as the visualization results of YOLOv8n, and (c1), (c2) and (c3) are denoted as the visualization results of DGSP-YOLO.

In summary, this study employed heatmap visualization techniques to compare the performance of the DGSP-YOLO model with the traditional YOLOv8 model in terms of feature extraction and target detection. Our analysis indicates that DGSP-YOLO exhibits significant advantages in several key aspects:

1) Precise Feature Focus: The DGSP-YOLO model is capable of concentrating more intently on features directly related to the target, such as highlighting the hull of a ship in maritime samples. In contrast, the YOLOv8 model tends to focus more on environmental features that are not directly associated with the ship.
2) Detection Capability for Small-Scale Targets: In datasets that include small-scale targets, the DGSP-YOLO model demonstrates superior performance, with its heatmap consistently covering the entire target area. This indicates a stronger ability to capture information

from small-scale targets, thereby enhancing recognition accuracy.

3) Adaptability to Target Size Variations: The DGSP-YOLO model effectively focuses on ship targets of varying sizes rather than irrelevant details, emphasizing its excellent performance in detecting ships in Synthetic Aperture Radar (SAR) images.

These advantages suggest that the DGSP-YOLO model significantly outperforms traditional methods in terms of accuracy and robustness in target detection, especially in complex environments for detecting small-scale targets.

V. CONCLUSION

This paper introduces a pioneering Synthetic Aperture Radar (SAR) ship target detection model, designated DGSP-YOLO. The model has been meticulously engineered to deliver superior performance across a spectrum of environmental conditions, ensuring precise detection and localization of maritime vessels. Extensive experimental validation using ship datasets, including HRSID, LS-SSDD-v1.0, and SSDD, has established DGSP-YOLO's marked superiority over existing models such as SSD, Faster R-CNN, RTDETR, and the YOLO series. Notably, DGSP-YOLO has achieved mAP50 metrics of 94%, 69%, and 99% on the HRSID, LS-SSDD-v1.0, and SSDD datasets, respectively. These results attest to the model's exceptional efficiency and robust generalization capabilities.

It is particularly noteworthy that DGSP-YOLO has realized significant enhancements in two pivotal performance metrics compared to the benchmark model: accuracy and recall. The model's performance in SAR ship detection is impressive, adeptly handling a variety of complex environmental conditions and ship target detection tasks across different scales.

However, we acknowledge that optimizing real-time performance and efficiency for practical applications remains a critical area of ongoing research. The current model faces challenges in achieving adequate detection accuracy within resource-constrained deployment scenarios. To overcome these challenges, we intend to pursue further model optimization in future research endeavors, aiming to develop a more streamlined design better aligned with real-world industrial requirements.

REFERENCES

[1] B. Wei, H. Li, T. Zhou, and S. Xing, "Obtaining 3D high-resolution underwater acoustic images by synthesizing virtual aperture on the 2D transducer array of multibeam echo sounder," *Remote Sens.*, vol. 11, no. 22, p. 2615, Nov. 2019.

[2] H.-M. Choi, H.-S. Yang, and W.-J. Seong, "Compressive underwater sonar imaging with synthetic aperture processing," *Remote Sens.*, vol. 13, no. 10, p. 1924, May 2021.

[3] X. Zhang, P. Yang, Y. Wang, W. Shen, J. Yang, J. Wang, K. Ye, M. Zhou, and H. Sun, "A novel multireceiver SAS RD processor," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4203611.

[4] X. Zhang, P. Yang, Y. Wang, W. Shen, J. Yang, K. Ye, M. Zhou, and H. Sun, "LBF-based CS algorithm for multireceiver SAS," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, Art. no. 1502505.

[5] W. Stecz and K. Gromada, "UAV mission planning with SAR application," *Sensors*, vol. 20, no. 4, p. 1080, Feb. 2020.

[6] J. Li, C. Xu, H. Su, L. Gao, and T. Wang, "Deep learning for SAR ship detection: Past, present and future," *Remote Sens.*, vol. 14, no. 11, p. 2712, Jun. 2022.

[7] X. Tan, X. Leng, R. Luo, Z. Sun, K. Ji, and G. Kuang, "YOLO-RC: SAR ship detection guided by characteristics of range-compressed domain," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 17, pp. 18834–18851, 2024.

[8] Z. Sun, X. Leng, X. Zhang, B. Xiong, K. Ji, and G. Kuang, "Ship recognition for complex SAR images via dual-branch transformer fusion network," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.

[9] X. Zhang, S. Feng, C. Zhao, Z. Sun, S. Zhang, and K. Ji, "MGSFA-Net: Multiscale global scattering feature association network for SAR ship target recognition," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 17, pp. 4611–4625, 2024.

[10] X. Tan, X. Leng, Z. Sun, R. Luo, K. Ji, and G. Kuang, "Lightweight ship detection network for SAR range-compressed domain," *Remote Sens.*, vol. 16, no. 17, p. 3284, Sep. 2024.

[11] Z. Zhou, L. Zhao, K. Ji, and G. Kuang, "A domain-adaptive few-shot SAR ship detection algorithm driven by the latent similarity between optical and SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5216318.

[12] M. A. Hussain, Z. Chen, M. Shoaib, S. U. Shah, J. Khan, and Z. Ying, "Sentinel-1A for monitoring land subsidence of coastal city of Pakistan using persistent scatterers in-SAR technique," *Sci. Rep.*, vol. 12, no. 1, p. 5294, Mar. 2022.

[13] A. Kshetrimayum, A. Goyal, and B. K. Bhadra, "Semi physical and machine learning approach for yield estimation of pearl millet crop using SAR and optical data products," *J. Spatial Sci.*, vol. 69, no. 2, pp. 573–592, Apr. 2024.

[14] X. Leng, K. Ji, K. Yang, and H. Zou, "A bilateral CFAR algorithm for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1536–1540, Jul. 2015.

[15] N. Li, X. Pan, L. Yang, Z. Huang, Z. Wu, and G. Zheng, "Adaptive CFAR method for SAR ship detection using intensity and texture feature fusion attention contrast mechanism," *Sensors*, vol. 22, no. 21, p. 8116, Oct. 2022.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, May 2020.

[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[18] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[20] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands. Cham, Switzerland: Springer, Oct. 2016, pp. 21–37.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[22] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery," *Remote Sens.*, vol. 11, no. 5, p. 531, Mar. 2019.

[23] H. Guo, X. Yang, N. Wang, and X. Gao, "A CenterNet++ model for ship detection in SAR images," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107787.

[24] W. Bao, M. Huang, Y. Zhang, Y. Xu, X. Liu, and X. Xiang, "Boosting ship detection in SAR images with complementary pretraining techniques," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 14, pp. 8941–8954, 2021.

[25] X. Wang, W. Xu, P. Huang, and W. Tan, "MSSD-Net: Multi-scale SAR ship detection network," *Remote Sens.*, vol. 16, no. 12, p. 2233, Jun. 2024.

[26] G. Tang, Y. Zhuge, C. Claramunt, and S. Men, "N-YOLO: A SAR ship detection using noise-classifying and complete-target extraction," *Remote Sens.*, vol. 13, no. 5, p. 871, Feb. 2021.

[27] J. Jiang, X. Fu, R. Qin, X. Wang, and Z. Ma, "High-speed lightweight ship detection algorithm based on YOLO-V4 for three-channels RGB SAR image," *Remote Sens.*, vol. 13, no. 10, p. 1909, May 2021.

[28] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[29] Z. Sun, X. Leng, Y. Lei, B. Xiong, K. Ji, and G. Kuang, "BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images," *Remote Sens.*, vol. 13, no. 21, p. 4209, Oct. 2021.

[30] S. Wang, S. Gao, L. Zhou, R. Liu, H. Zhang, J. Liu, Y. Jia, and J. Qian, "YOLO-SD: Small ship detection in SAR images by multi-scale convolution and feature transformer module," *Remote Sens.*, vol. 14, no. 20, p. 5268, Oct. 2022.

[31] Y. Guo, S. Chen, R. Zhan, W. Wang, and J. Zhang, "LMSD-YOLO: A lightweight YOLO algorithm for multi-scale SAR ship detection," *Remote Sens.*, vol. 14, no. 19, p. 4801, Sep. 2022.

[32] X. Ren, Y. Bai, G. Liu, and P. Zhang, "YOLO-lite: An efficient lightweight network for SAR ship detection," *Remote Sens.*, vol. 15, no. 15, p. 3771, Jul. 2023.

[33] Z. Liangjun, N. Feng, X. Yubin, L. Gang, H. Zhongliang, and Z. Yuanyang, "MSFA-YOLO: A multi-scale SAR ship detection algorithm based on fused attention," *IEEE Access*, vol. 12, pp. 24554–24568, 2024.

[34] R. Varghese and M. Sambath, "YOLOv8: A novel object detection algorithm with enhanced performance and robustness," in *Proc. Int. Conf. Adv. Data Eng. Intell. Comput. Syst. (ADICS)*, Apr. 2024, pp. 1–6.

[35] X. Tang, J. Zhang, Y. Xia, and H. Xiao, "DBW-YOLO: A high-precision SAR ship detection method for complex environments," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 17, pp. 7029–7039, 2024.

[36] R. Sunkara and T. Luo, "No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, Sep. 2022, pp. 443–459.

[37] A. Srinivas, T.-Y. Lin, N. Parmar, J. Shlens, P. Abbeel, and A. Vaswani, "Bottleneck transformers for visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16514–16524.

[38] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1580–1589.

[39] W. Liu, H. Lu, H. Fu, and Z. Cao, "Learning to upsample by learning to sample," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 6027–6037.

[40] J. Wang, K. Chen, R. Xu, Z. Liu, C. Change Loy, and D. Lin, "CARAFE: Content-aware ReAssembly of FEatures," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3007–3016.

[41] H. Lu, W. Liu, H. Fu, and Z. Cao, "Fade: Fusing the assets of decoder and encoder for task-agnostic upsampling," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 231–247.

[42] H. Lu, W. Liu, Z. Ye, H. Fu, Y. Liu, and Z. Cao, "SAPA: Similarity-aware point affiliation for feature upsampling," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 20889–20901.

[43] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient IOU loss for accurate bounding box regression," *Neurocomputing*, vol. 506, pp. 146–157, Sep. 2022.

[44] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 658–666.

[45] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance—IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2020, vol. 34, no. 7, pp. 12993–13000.

[46] H. Zhang and S. Zhang, "Shape-IoU: More accurate metric considering bounding box shape and scale," 2023, *arXiv:2312.17663*.

[47] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.

[48] T. Zhang, X. Zhang, X. Ke, X. Zhan, J. Shi, S. Wei, D. Pan, J. Li, H. Su, Y. Zhou, and D. Kumar, "LS-SSDD-v1.0: A deep learning dataset dedicated to small ship detection from large-scale Sentinel-1 SAR images," *Remote Sens.*, vol. 12, no. 18, p. 2997, Sep. 2020.

[49] T. Zhang, X. Zhang, J. Li, X. Xu, B. Wang, X. Zhan, Y. Xu, X. Ke, T. Zeng, H. Su, I. Ahmad, D. Pan, C. Liu, Y. Zhou, J. Shi, and S. Wei, "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sens.*, vol. 13, no. 18, p. 3690, Sep. 2021.

[50] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13713–13722.

[51] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 7132–7141.

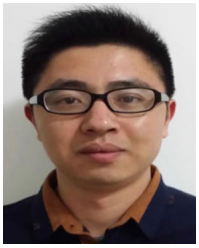
[52] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11534–11542.

[53] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.

[54] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs beat YOLOs on real-time object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 16965–16974.

[55] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

ZHU LEJUN was born in Fuyang, China, in 2002. He received the bachelor's degree in Internet of Things engineering from Anhui Sanlian University, in 2023. He is currently pursuing the master's degree in computer science with Hubei University of Technology. His research interests include image recognition and deep learning.



CHEN JINGLIANG received the Ph.D. degree in computer software and theory from Wuhan University, Hubei, China, in 2015. He is currently with the School of Computer Science, Hubei University of Technology. His current research interests include software engineering, intelligent computing, and image segmentation.

CHEN JIAYU was born in Ezhou, Hubei, China, in 2000. He received the bachelor's degree in software engineering from Hubei University of Technology, in 2022, where he is currently pursuing the master's degree in computer science and technology. His research interests include data analysis and anomaly detection and machine learning.

YANG HAO was born in Yichun, China, in 1999. He received the bachelor's degree in civil engineering from Jiangxi Institute of Applied Science and Technology, in 2023. He is currently pursuing the master's degree in computer science with Hubei University of Technology. His research interests include image recognition and deep learning.