A Method for Nearshore Vessel Target Detection in SAR Imagery Utilizing Edge Characteristics and Augmented Global Information Amplification

Hongjian Ye, Weiming Chen, Diyong Wang, and Yu Qiu

***Abstract*—Synthetic aperture radar (SAR), which can work normally under various meteorological conditions, has been widely researched and applied in marine vessel target monitoring and identification. Among the many research topics, due to the inconsistency of ship scale in SAR images, susceptibility to sea and land noise and clutter interference, resulting in a low detection rate of near-shore ship targets and inaccurate edge delineation of densely lined ships, a target detection algorithm based on deep convolutional neural network is proposed. The algorithm employs the channel-space grouping attention mechanism during feature extraction to enhance features by utilizing global positional and edge information associated with instances. The feature mobility fusion module is employed to merge features of various scales, bolster the interconnection among these features, and enhance multiscale ship target detection capabilities. The decoupled head is employed for ship target localization, while the angle-weighted intersection over union is used to mitigate regression errors. The experimental results show that the precision (P) achieved on HRSID and SSDD datasets reaches 94.81% and 99.01%, respectively, exceeding the control algorithm by more than 1.35% and 0.94%; the average precision (mAP) reaches 92.06% and 99.50%, respectively, exceeding the control algorithm by more than 2.32% and 2.51%; this indicates that the proposed algorithm has a good performance on SAR image ship detection and a strong generalization ability.**

***Index Terms*—Anchorless box mechanism, channel-space grouping attention mechanisms (CSG), feature mobility fusion (FMF) module, synthetic aperture radar (SAR).**

I. Introduction

MARINE vessel monitoring and identification are vital for coastal nations, aiding in the protection of marine resources, fisheries vessel oversight, and the prevention of smuggling and stowaways. Synthetic aperture radar (SAR) technology provides all-weather active imaging capabilities, unaffected by weather and lighting conditions [1], [2]. With the continuous advancements in SAR imaging technology, image resolution continues to improve [3], [4]. In recent years, substantial research has focused on the effective extraction and utilization of SAR image data for marine vessel monitoring and identification.

Traditional SAR image ship target detection algorithms rely on contrast disparities between targets and background clutter and employ mathematical and statistical methods, such as constant false-alarm rate (CFAR) detection, template matching, and wake modeling [5], [6], [7], [8], [9]. Pan et al. [10] incorporated a generalized gamma distribution clutter model into the CFAR algorithm for multiscale target detection, considering practical scenarios, although the approach exhibits limited generalizability. Leng et al. [11] introduced bilateral CFAR to mitigate the impact of SAR image blurriness and background noise on detection. The aforementioned methods rely on manually crafted features, resulting in low detection efficiency and limited generalization capabilities, making them unsuitable for ship target detection in complex environments.

The rapid development of deep learning in computer vision [12], [13], [14], [15], [16], [17], [18], [19] has led to the emergence of various target detection approaches, including single-stage and two-stage detection, anchor box detection, anchorless box detection, and the use of transformers in SAR image detection applications. Deng et al. [20] utilized a dense network to repurpose feature information, but it is less suitable for densely arranged ships. Li et al. [21] improved feature extraction by altering the convolution kernel's sampling position. Sun et al. [22] based on the YOLO detection framework, solved the problem of discontinuous ship boundaries by designing an angle classification module to capture feature target length, width, and direction information. Humayun et al. [23] formulated anchor frames matching the targets based on the data distribution of target sizes in the dataset to enhance the network's sensitivity to targets of specific sizes and to improve the training efficiency and detection accuracy, but it is easy to ignore targets outside the specific size interval. Anchor-less box detection has gained widespread adoption for accurately delineating ship boundaries in SAR image target detection [24], [25], [26], [27], [28], [29], [30]. Sun et al. [31] performed pixel-by-pixel prediction of images to reduce false positives and misses based on the anchorless frame mechanism and optimized the target location using the CP module. Du et al. [32] introduced reinforcement learning

Manuscript received 27 December 2023; revised 24 March 2024 and 10 April 2024; accepted 11 April 2024. Date of publication 16 April 2024; date of current version 29 April 2024. This work was supported in part by the National Key R&D Program of China under Grant 2022YFC3005900, in part by the Fundamental Research Funds for the Central Universities, China University of Geosciences under Grant CUG2642022006, and in part by the Open Fund Project of State Key Laboratory of Gas Disaster Detecting, Preventing and Emergency Controlling under Grant 2022SKLKF08. *(Corresponding author: Weiming Chen.)*

Hongjian Ye, Diyong Wang, and Yu Qiu are with the Faculty of Engineering, China University of Geosciences, Wuhan 430074, China (e-mail: yehongjian@cug.edu.cn; diyongwang@hotmail.com; qiuyu@cug.edu.cn).

Weiming Chen is with the Faculty of Engineering and Institute for Natural Disaster Risk Prevention and Emergency Management, China University of Geosciences, Wuhan 430074, China (e-mail: chenweiming@cug.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3389210

to autonomously identify ship-containing regions and mitigate background noise influence. Hou et al. [33] implemented scene interpretation before detection to reduce the impact of background noise.

Enhancing the quality of semantic information significantly improves SAR image detection performance, with attention mechanisms and feature pyramid networks (FPNs) widely adopted [34], [35]. Bai et al. [36] improved the detection accuracy by aligning spatial attention and target features into a feature pyramid with shallow feature reorganization to reduce scattering noise and obtain rich semantic information. Regarding attention mechanisms, Chen and Gao [37] introduced an attention transfer mechanism, masking attention for all ship targets to clarify semantic feature information. Feng et al. [38] proposed position-enhanced attention by capturing horizontal and vertical ship attention maps and fusing them with feature maps to enrich semantic information. Concerning FPN fusion, Bai et al. [36] introduced a feature enhancement pyramid, enhancing features through a spatial enhancement module and feature alignment module by combining deep and shallow semantic features. Zhang et al. [39] utilized FPN to merge features from the backbone network at different scales, enhancing attention to small targets.

Many existing methods rely on spatial or convolutional block attention mechanisms, making networks overly focused on local target details, akin to RAM [40] and CBAM [41]. This hyper-local focus can result in semantic information being obscured by background noise. Directly fusing feature information containing complex background noise can exacerbate semantic ambiguity issues. Current research often emphasizes deepening feature fusion without fundamentally addressing semantic ambiguity problems. Current SAR image ship target detection methods predominantly rely on horizontal anchor frame algorithms [42], [43], [44]; directional anchor frame algorithms [45], [46], [47]; and anchor frame-free detection algorithms [24], [25], [26], [27], [28], [29], [30]. However, the introduction of anchor frames can consume substantial computational resources due to the discontinuous boundaries of SAR image instances, varying instance density, and scattering noise. While anchorless frame detection algorithms overcome the anchor frame limitations, they mainly address regression issues using center bias distance loss, leading to significant regression errors. The challenges in the above-mentioned SAR image detection task are summarized as follows.

1) Existing networks are overly concerned with the local details of the target, and due to various scattering phenomena resulting in discontinuous, noisy, or deformed object boundaries, it is very easy to generate false alarms due to the fact that the image edges can provide little effective semantic information, and there may be features that are similar to the target information.
2) Ship detection in SAR images near the shore presents challenges due to complex backgrounds, such as waves, islands, and harbors. Some features or noise from these backgrounds may resemble ship features, resulting in compromised detection accuracy while aiming for high recall rates.

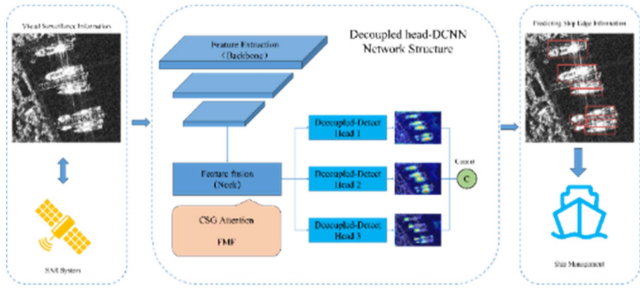

Fig. 1. Flowchart of SAR image ship detection.

3) In contrast to optical images, SAR images lack effective RGB information, leading to semantic ambiguity issues.
4) The prevalence of ocean areas in the images and the absence of labeled data for land in the current dataset make it challenging to learn features for small ships.

To address the above-mentioned challenges, a network model of "decoupled head-deep convolutional neural network (DCNN)" is proposed. Our approach utilizes the CSPDarknet53 framework for initial feature extraction, and uses the channel-space grouping attention mechanisms (CSG) module, and the feature mobility fusion (FMF) module in the feature fusion network to extract and fuse the feature information, to obtain a feature map containing rich semantic information. In addition, an angle-based weighted intersection over union (angle-WIoU) loss function is introduced to correct the detection loss of the decoupled detection head based on the angle between the line connecting the center of attention of the true and predicted values and the horizontal direction.

The primary research findings and contributions are outlined as follows.

1) A CSG is designed to extract feature maps containing high-quality semantic information by enhancing the grouping characteristics of channel and spatial features, thereby enhancing the network's ability to discriminate between noise and target features and improving the detection performance of targets at different scales.
2) Fusing the input feature maps using the FMF module, which facilitates interscale information interaction, makes the network applicable to ship target detection at different scales, and reduces false alarms at the edges of images.
3) An angular-WIoU loss function is developed for the anchorless box detection head to minimize object loss, and nonmonotonic focusing coefficients are employed to mitigate gradient degradation caused by distance-corrected and angular-corrected regressions for low-quality samples and accurately locate the center of the target.
4) The proposed algorithm's practicality, accuracy, and generalizability for detecting near-shore ship targets in SAR images are showcased through comparative experiments on various datasets.

The rest of this article is organized as follows. In Section II, the proposed method is summarized. In Section III, experimental details are given. In Section IV, the experimental results are analyzed. In Section V, the experimental results and analysis

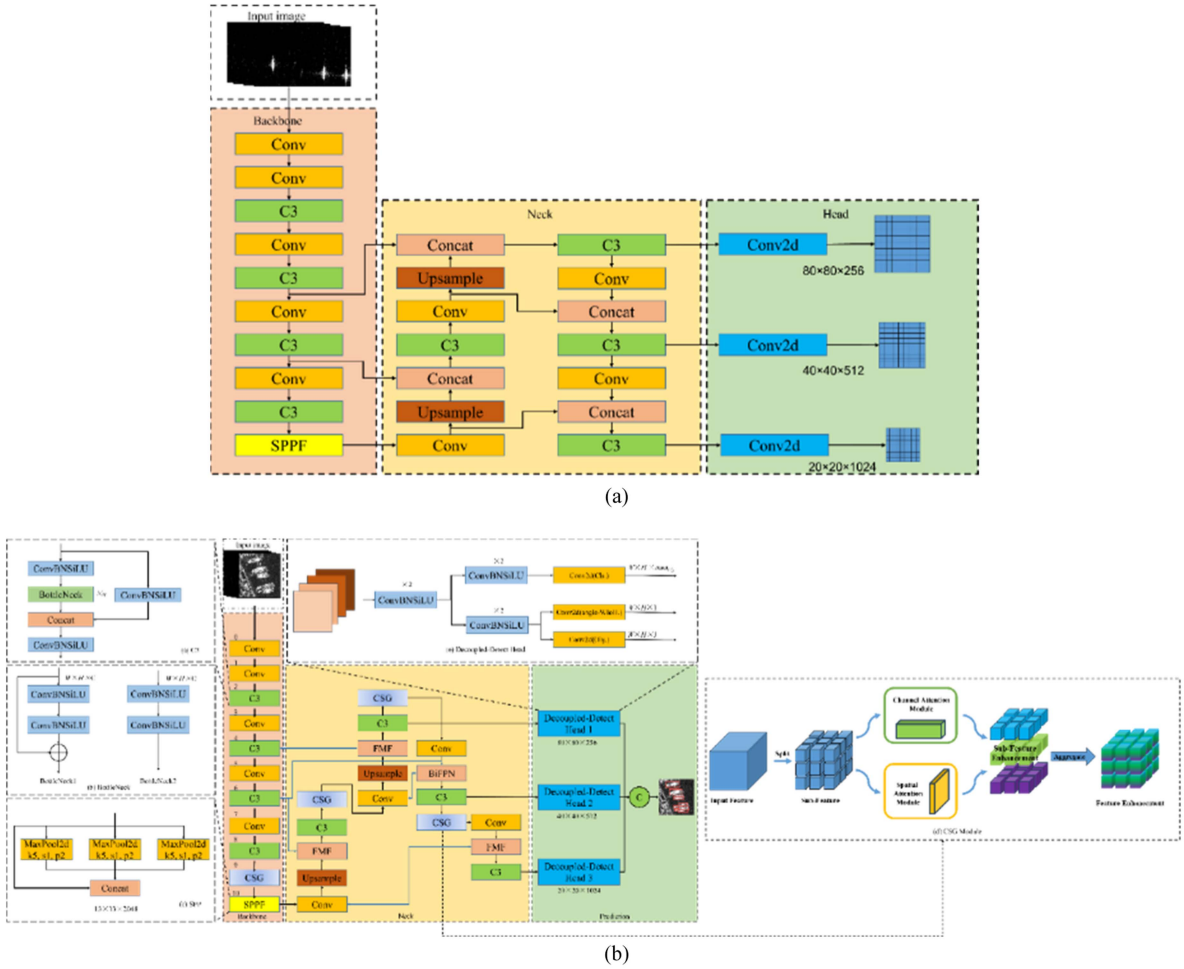

Fig. 2. (a) Schematic diagram of YOLOv5 network structure. (b) Schematic diagram of decoupled head-DCNN network structure.

are discussed. Finally, in Section VI, this article is summarized and limitations of this study and the future research directions are pointed out.

II. Proposed Method

Fig. 1 depicts the flow of the decoupled head-DCNN algorithm, consisting of three elements: SAR image input, the decoupled head-DCNN network for image feature extraction and depth information fusion, and ship prediction results.

A. Decoupled Head-DCNN Network Architecture

The proposed approach is a single-stage detection method, and its network structure in the feature extraction stage is inspired by YOLOv5, in Fig. 2(a). The proposed model is depicted in Fig. 2(b), and an in-depth explanation will be provided in the subsequent sections.

1) Backbone: The CSPDarknet53 network serves as the feature extraction backbone through Conv+C3. The output feature maps from the last three C3 modules undergo feature enhancement, with output image sizes at 1/8, 1/16, and 1/32 of the input image scale. The final feature fusion is accomplished through the feature SPPF module.

2) Feature Enhancement: To improve multiscale ship detection performance, the features extracted by the backbone are enhanced. In the Neck stage, the introduced CSG attention module enhances the attention for position and fine features while mitigating background noise interference, which includes elements like docks, buildings, islands, and ship wakes. The FMF process, as represented by the blue arrows in Fig. 2(b), is facilitated by the FMF module, combining both strong and weak semantic information to generate rich semantic features.

3) Decoupled Head: This method utilizes a multiscale detection feature fusion strategy for high-precision detection of arbitrarily oriented ships in SAR images. The decoupled head, as illustrated in Fig. 2(b), comprises three branches: *Cls.* for class prediction loss, *Obj.* for target presence probability prediction loss, and *angle-WIoU.* for bounding box prediction loss.

B. Channel-Space Grouping Attention Mechanism

To mitigate the impact of sea and land noise in SAR images, inspired by the concept introduced by Zhang and Yang [48], we propose the CSG attention module, as depicted in Fig. 3. This module is positioned before the SPPF module of the backbone and after the upsampling in the Neck. It efficiently processes

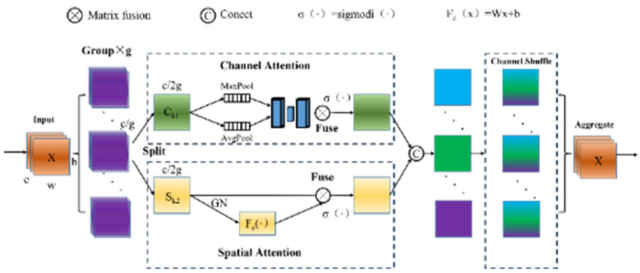

Fig. 3. Schematic diagram of CSG attention.

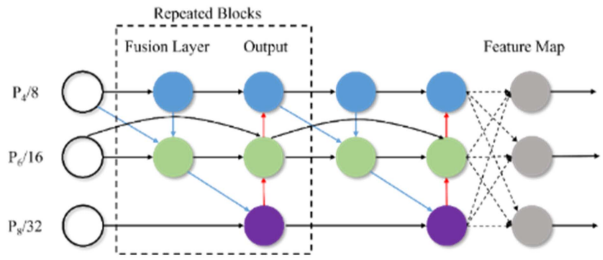

Fig. 4. Schematic diagram of BiFPN structure.

local and global feature information, thereby improving global localization accuracy and instance segmentation.

1) Grouping: Given a feature mapping $X \in R^{C\times H\times W}$, where *C*, *H*, and *W* denote the number of channels, height, and width, respectively, the feature *X* is divided into group *g*. (In this article, *g* is set to 8.) The attention module generates corresponding weight coefficients for each subfeature in accordance with the degree of importance, and each group captures a specific semantic feature, and channel and spatial attention enhancement is carried out along the channel dimension, respectively.

The attention calculation process divides the feature channel into eight different groups representing different features learned and calculates the respective attention weights within each group, which allows the channel features within the group to exchange information and avoids the mutual influence of the features in different groups, which in turn strengthens the learning ability of the attention mechanism for the feature correlation, and enhances the differentiation ability between the feature and the nonfeature elements.

2) Channeling Attention: Each group of feature information is fed into two parallel operations, average pooling (*AvgPool*) and maximum pooling (*MaxPool*), to incorporate global information and produce channel statistics, $s \in R^{\frac{c}{2g}\times 1\times 1}$. This process compresses the global spatial dimensions $H \times W$, resulting in the creation of compressed channel features denoted as X_{k1}, this utilization of global information enhances target localization capability. In (1), X_{k1} represents the channel features, while H, W represent the spatial dimensions.

$$s = F_{gp}(X_{k1}) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{k1}(I,J). \tag{1}$$

The pooled outputs are element-wise summed and then input into a two-layer neural network designed for information sharing. In the first layer of this network, there are *channel/r* neurons (where r is the reduction rate), utilizing a *ReLU* activation function. The second layer consists of neurons organized in a channel format. Through a summation operation and the application of a sigmoid (σ) activation function, a compact feature channel, referred to as X_{k1}', is created. This principle is illustrated in the following equations:

$$MPL(s) = \sigma(conv(\mathrm{Re}LU(conv(s)))) \tag{2}$$

$$M_c(F) = MLP(AvgPool(F_c(s)) + MLP(MaxPool(F_c(s)) \tag{3}$$

$$X_{k1}{}' = \sigma(W_1(M_c(F)) + b_1). \tag{4}$$

3) Spatial Attention: The CSG method leverages group norm to capture spatial statistics during spatial attention processing. This aids in emphasizing feature maps with valuable information content while reducing the impact of noise and irrelevant data. Feature weighting is employed to produce enhanced spatial features denoted as $X_{k2}{}'$, ultimately boosting the network's ability for target segmentation. These enhanced spatial features $X_{k2}{}'$ are determined in the following equation:

$$X_{k2}{}' = \sigma(W_2 \cdot GN(X_{k2}) + b_2) \cdot X_{k2}. \tag{5}$$

4) Feature Aggregation: The enhanced features are integrated during the feature aggregation stage. The "channel shuffle" method is used for feature fusion [34], in which each group of fused features strengthened by channel and spatial attention is divided into g subgroups, which are exchanged to form a new feature and fed into the next layer for aggregation, realizing cross-group information interaction along the channel dimension.

C. FMF Module

Compared with the traditional FPN network, inspired by the concept of BiFPN, jump connections are introduced between input and output features at the same layer. This facilitates the efficient transfer and fusion of feature information across scales in both directions, promoting the effective transmission of elemental information [49]. The structure of BiFPN is schematically depicted in Fig. 4, with blue arrows representing the top-down pathway for high-level semantic information transmission, the red portion denoting the bottom-up transfer of underlying semantic information, and the black arrows signifying connections and jump connections within the same layers.

The FMF module integrates weighted individual scale features using a multistep process. Initially, the network autonomously learns weight parameters for each input layer, enabling a two-way information flow interaction across different scales. Subsequently, the model conducts feature fusion by placing the feature information of adjacent scales in the middle layer, promoting bottom-up feature integration while retaining fine-grained details globally through maximum pooling. Lastly, the fused feature maps undergo top-down fusion, preserving global texture information with average pooling. Finally, the

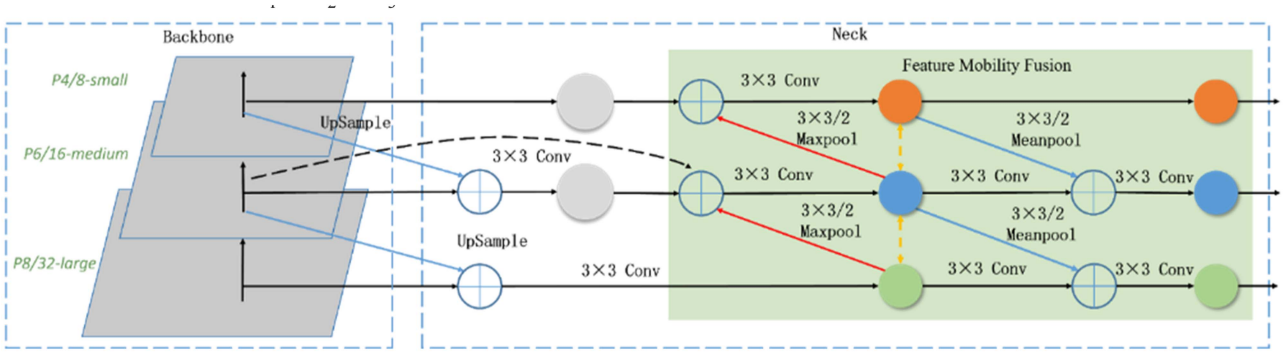

Fig. 5. Schematic of FMF multiscale feature fusion.

processed features are ready for output. Equations (6) and (7) are top-down feature fusion and (8) is used to aggregate multiscale features.

$$P_i^{\text{td}} = Conv\left(\frac{\omega_1 \cdot P_i^{in} + \omega_2 \cdot {P_{i+1}}^{in}}{\omega_1 + \omega_2 + \varepsilon}\right) \tag{6}$$

$$P_i^{out} = Conv\left(\frac{\omega_1' \cdot P_i^{in} + \omega_2' \cdot P_i^{td} + \omega_3' \cdot {P_{i-1}}^{\text{out}}}{\omega_1' + \omega_2' + \omega_3' + \varepsilon}\right) \tag{7}$$

$$P_i^{out} = Conv(P_i^{in} + \text{Resize}({P_{i+1}}^{Out})). \tag{8}$$

In the above-mentioned equation, P_i^{td} represents the intermediate feature in the top-down path, while P_i^{out} denotes the output feature of the *i*th layer from top to bottom. ω_1, ω_2 corresponds to the weights of the current input layer, while $\omega_1', \omega_2', \omega_3'$ are the weight parameters of the output layer. ε denotes the correction parameter, and *Resize* indicates the operation of upsampling or subsampling.

The FMF module is applied to perform feature fusion in the Neck section of the model, as shown in Fig. 5, "+" denotes the fusion of features from different scales, the gray circle represents the middle layer, and the colored circles represent the output layers corresponding to various scales. Yellow arrows are used to depict feature interactions among these output layers. Within the Neck, the repeated blocks process, as illustrated in Fig. 4, is repeatedly executed to optimize weighted features and enhance the clarity of feature importance. This iterative learning process in the Neck section also involves the introduction of yellow arrows to signify the top-down transfer of high-level semantic information. By enabling the flow of image feature information across different scales, this methodology mitigates semantic ambiguities and enhances result accuracy.

D. Angle-WIoU Loss Function

The chosen detection head is illustrated in Fig. 6, with three branches serving the functions of classification, refinement, and regression. Specifically, the *Cls*. branch employs binary cross-entropy loss to measure the disparity between the predicted category and the actual category. The *Obj*. branch utilizes binary cross-entropy loss to evaluate the deviation between the predicted probability of target presence and the true probability of its existence. Finally, the *angle-WIoU*. branch calculates

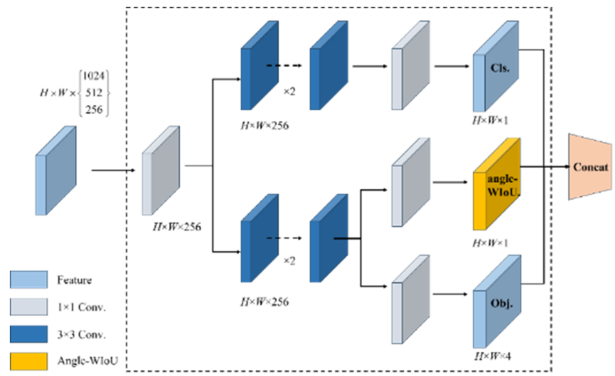

Fig. 6. Schematic diagram of decoupled head structure.

the difference in cross-combination ratio between the model's predicted location and the actual location.

The decoupled detection head can combine different target localization modules more flexibly, and can better balance the two tasks of target localization and target classification, reduce the risk of overfitting, reduce the complexity of the model, and enable the model to be optimized more easily. In addition, the decoupled detection header enables different parts of the network to carry out migration learning independently, and the model trained on one dataset can be applied to more similar models, so the proposed model adopts the decoupled detection header.

The decoupled head conducts feature map regression, as illustrated in Fig. 7(a). It employs the geometric center point of the feature map as the reference and defines an a priori prediction region with a 2.5 times step size. Subsequently, it quantifies the disparity between the a priori prediction frames and the ground truth using regression correction via angle-WIoU, as demonstrated in Fig. 7(b). The final output comprises the regression results.

To mitigate the influence of wave noise and scattering noise inherent in SAR image samples, a precise regression penalty loss is introduced. This loss function, called the angle-WIoU loss, is designed specifically for bounding box regression. It incorporates an angle penalty term, which is constructed based on WIoU. Additionally, the angle penalty term is combined with a nonmonotonic focusing mechanism to mitigate the adverse effects of low-quality instances on the regression process.

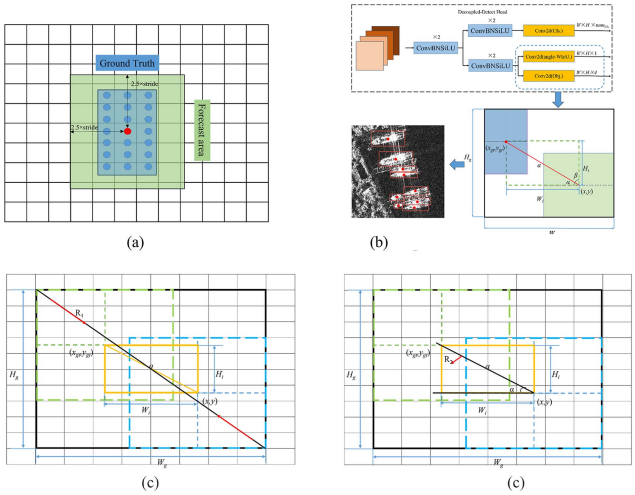

Fig. 7. Schematic diagram of ship target detection regression. (a) Target regression for the normal case. (b) Target regression for correcting low-quality samples using the angle-WIoU loss function. The green region is the prediction region framed by the feature map, and the blue region is ground truth. (c) Distance correction. (d) Angle correction. (a) Target regression for ships. (b) Ship target regression correction. (c) R1 penalty. (d) R2 Penalty.

1) Distance Correction: The distance penalty term is constructed as shown in Fig. 7(c), and the distance correction is performed using the line connecting the centroids of the true and calculated values. The distance penalty term is defined in the following equation:

$$R_1 = \exp\left(\frac{(x_i - x_{gt})^2 + (y_i - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \tag{9}$$

where R_1 represents the distance penalty term, and "$*$" denotes that W_g and H_g have been detached from the computational graph to prevent interference with gradient descent convergence.

2) Angle Correction: Formulate the angle penalty term, illustrated in Fig. 7(d), which employs the angle between the line connecting the true and predicted centroids and the horizontal direction for angle adjustment. The angle penalty term is specified in (14).

$$\begin{aligned}\Delta &= 1 - 2\sin^2(\arcsin(H_i/a) - \pi/4) \\ &= \cos(2(\arcsin(H_i/a) - \pi/4))\end{aligned} \tag{10}$$

$$a = \sqrt{(x_{gt} - x)^2 + (y_{gt} - y)^2} \tag{11}$$

where a is the distance between the centroids, and $\frac{H_i}{a}$ denotes the sine of the angle between the line connecting the centroids and the horizontal direction $\sin\alpha$.

$$H_i = \max(y_{gt}, y) - \min(y_{gt}, y) \tag{12}$$

$$R_2 = \frac{1}{2}\sum_{n=x,y}\left(1 - \exp\left(-\gamma\left(\frac{n_{gt} - n}{\rho_n}\right)^2\right)\right), \begin{cases}\rho_x = W_g \\ \rho_y = H_g\end{cases} \tag{13}$$

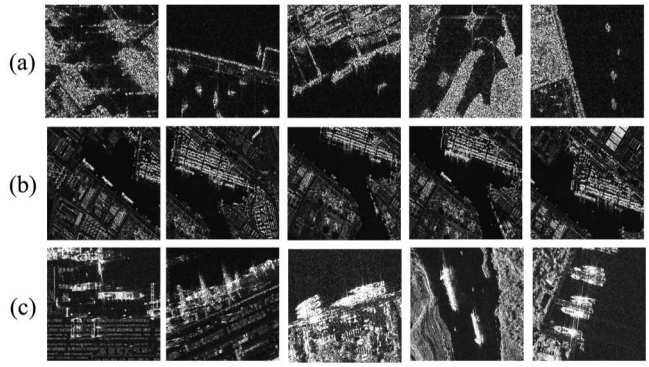

Fig. 8. Experimental sample display. (a) Sample in the SAR-Ship-Dataset. (b) Sample in the HRSID dataset. (c) Sample in the SSDD dataset.

$$R_2 = 1 - \frac{\exp\left(-\gamma\left(\frac{x_{gt}-x}{W_g}\right)^2\right) + \exp\left(-\gamma\left(\frac{y_{gt}-y}{H_g}\right)^2\right)}{2}, \gamma = 2 - \Delta \tag{14}$$

where R_2 is the angular penalty term, and Δ denotes the angular correction term.

3) Boundary Box Loss: Building upon the previous concepts, an outlier degree is introduced to quantify the quality of instances. Smaller outlier degrees correspond to higher-quality instances, and targets with larger outlier degrees are assigned smaller gradient gains to mitigate the impact of low-quality instances, as defined in (15).

$$\beta = \frac{L^*_{IoU}}{\overline{L_{IoU}}} \in [0, +\infty) \tag{15}$$

$$L_{\text{angle-WIoU}} = rL_{IoU}(R_1 + R_2), r = \frac{\beta}{\delta\lambda^{\beta-\delta}}. \tag{16}$$

The ultimate loss function is presented in (16), where L_{IoU} is the *IoU* loss; L^*_{IoU} denotes the gradient gain; $\overline{L_{IoU}}$ is the running mean of momentum m; λ and δ are hyperparameters that control the mapping of the nonmonotonic focusing coefficients β and the gradient gains r where λ is used to balance the weights of the intersection and concatenation sets, characterizing the importance given to correctly classified and misclassified regions, and δ is used to adjust the rate of change of the loss function, controlling the degree of smoothing of the loss function as the degree of overlap between the predicted and the true values changes. Both hyperparameters are determined by the validation set to determine their optimal values.

III. Experimental Details

A. Dataset and Experimental Preparation

Experiments were conducted using the SAR-Ship-Dataset, HRSID, and SSDD datasets. In the training process, the SAR-Ship-Dataset dataset is used for network weight training, and the HRSID dataset and SSDD dataset are used for network performance validation. The experimental samples are shown in Fig. 8. Where (a) is the sample in the SAR-Ship-Dataset; (b)

TABLE I
DETAILED DESCRIPTION OF SAR-SHIP-DATASET

Parameter	Value
Sensors	102 Chinese Gaofen-3 images, 108 Sentinel-1 images
Resolution(m)	3–25
Number of images	43 819
Number of instances	59 535
Image size	256 × 256
Environmental conditions	Complexity and simplicity
Scenes	Inshore, offshore

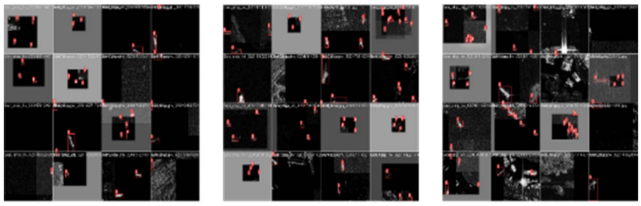

Fig. 9. Display of image enhancement results in SAR-Ship-Dataset.

TABLE II
DETAILED DESCRIPTION OF HRSID

Parameter	Value
Sensors	TerraSAR-X image, 1 TanDEM-X image
Resolution(m)	0.5–3
Number of images	5604
Number of instances	16951
Image size	800 × 800
Environmental conditions	Complexity
Scenes	Inshore, offshore

is the sample in the HRSID dataset; and (c) is the sample in the SSDD dataset.

1) The SAR-Ship-Dataset ([50]): The dataset comprises 102 Chinese Gaofen-3 images and 108 Sentinel-1 images, resulting in a total of 43 819 images and 59 535 instances. The images exhibit a range of target resolutions, varying from 3 to 25 m, with a fixed size of 256 × 256 pixels. Key parameters of the SAR-Ship-Dataset are detailed in Table I. To facilitate model training, we randomly partitioned the dataset into training, testing, and validation sets in a 7:2:1 ratio. Data augmentation techniques were applied, including rotation, scaling, panning, mosaic [51], and mixup [52]. An illustrative example of data augmentation can be found in Fig. 9.

2) The HRSID Dataset ([45]): The dataset consists of TerraSAR-X and TanDEM-X images, amounting to 5604 images and 16 951 instances. Target resolutions vary between 0.5 and 3 m, and the image size is 800 × 800 pixels. Essential parameters for HRSID can be found in Table II.

3) The SSDD Dataset ([46]): The dataset comprises RadarSat-2, TerraSAR, and Sentinel-1 images, totaling 1160 images and 2456 instances. Target resolutions vary from 1 to 15 m, and the image aspect size is approximately 600 pixels. Refer to Table III for the key parameters of SSDD.

The experiments were conducted on a Windows 11 system with an NVIDIA A100-PCIE-40 GB graphics card (GPU). The

TABLE III
DETAILED DESCRIPTION OF SSDD

Parameter	Value
Sensors	RadaeSat-2, TerraSAR, Sentinel-1
Resolution(m)	1–15
Number of images	1160
Number of instances	2456
Image size	Length and width approx. 600
Environmental conditions	Simplicity
Scenes	Inshore, offshore

TABLE IV
EXPERIMENTAL ENVIRONMENT AND HYPERPARAMETERS

Configure	Information
Hardware	Operating system: Windows11
	GPU: NVIDIA A100-PCIE-40 GB CPU: Intel(R) Xeon(R) Gold 6326 CPU @ 2.90 GHz RAM:256 G
Software	Python 3.9+Pytorch 1.12.0 cu114+CUDA 11.4+cuDNN8.2.4
Parameter	Batch size: 12
	Number of iterations: 300
	Learning rate: 0.0001
	Optimizer: Adam
	Scheduler: Cosine Annealing
	Input size: (640, 640)

model was implemented using the PyTorch framework and training was accelerated using CUDA and cuDNN. The accuracy may vary depending on the version of the computational framework. All experiments utilized a pretrained transformer model from the COCO dataset and employed a cosine annealing learning rate schedule. The optimizer used was Adam, with a total of 300 iterations. Optimizing hyperparameters can significantly enhance algorithm convergence. Considering a large amount of training data, to reduce the computational expenditure, to retain the integrity of the target features in the image during the training process, and to balance the computational efficiency and detection accuracy, the proposed method selects 640 × 640 size image as the network input size. Further details are provided in Table IV.

B. Evaluation Indicators

Evaluation metrics for this study include precision (P), recall (R), and average precision (AP), as defined in the following equations:

$$P = \frac{TP}{TP + FP} \tag{17}$$

$$R = \frac{TP}{TP + FN} \tag{18}$$

$$AP = \int_0^1 P(R)dR, mAP = \sum_{i=1}^{n} \frac{AP_i}{n}. \tag{19}$$

In the above-mentioned equation, P represents the ratio of correctly predicted targets to all predicted targets, while R denotes the proportion of actual positive samples correctly predicted. AP quantifies the area under the precision-recall (PR) curve, offering

TABLE V
ABLATION EXPERIMENTS

Num	Decoupled head	CSG	FMF	P (%)	R (%)	mAP_{medium} (%)	mAP_{small} (%)
1	×	×	×	0.871	0.874	0.896	0.483
2	×*	×	×	0.875	0.867	0.905	0.52
3	×	√	×	0.89	0.883	0.908	0.522
4	×	×	√	0.891	0.882	0.906	0.502
5	√*	×	×	0.884	0.881	0.904	0.554
6	√*	√	√	0.914	0.899	0.945	0.652

In Table V, "×" indicates no such module; "√" indicates the addition of the module; "*" indicates that the angle-WIoU loss function is used, and WIoU is used for the loss function of the control group; the control group of the decoupled head uses the anchor frame-based detection head.

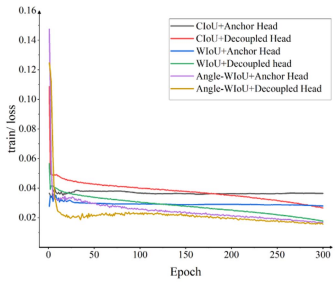

Fig. 10. Convergence of loss functions. The convergence effect is compared by training angle-WIoU, WIoU, and CIoU combined with different detection head mechanisms.

a comprehensive assessment of the model's detection performance. Higher *AP* values indicate better model performance. *mAP* represents the mean value for assessing the model's detection effectiveness across *n* scenes. Furthermore, the model's size was evaluated using params (*M*), and its detection speed was measured using latency (*ms*).

IV. Result and Analysis

A. Ablation Experiments

In the ablation experiments, groups 2 and 5 demonstrate the effectiveness of the angle-WIoU loss function combined with the decoupled head, while groups 3 and 4 highlight the efficacy of the CSG module and the FMF module, respectively. The corresponding experimental results are provided in Table V.

1) All the metrics of the proposed model were higher than the benchmark *P*, *R*, mAP_{medium}, and mAP_{small} reached 91.4%, 89.9%, 94.5%, and 65.2%, respectively, which is an enhancement of 4.3%, 2.5%, 4.9%, and 16.9%, respectively, compared with the control.
2) The introduction of angle-WIoU incorporates angle correction and nonmonotonic aggregation coefficients, specifically designed to address the challenge of precise localization for densely packed targets when using an anchorless frame detection head. The combined use of angle-WIoU and decoupled head yields the best detection results, addressing the issue of misclassifying low-quality small targets as background noise, thus reducing the omission and misdetection of small target ships near the shore. Loss convergence comparisons for various combinations of loss functions and detection heads are presented in Fig. 10, where the proposed method achieves a loss of approximately 0.015. The convergence results for the control groups with other combinations are higher than those of the proposed method, demonstrating the efficacy of the combination of angle-WIoU and decoupled head.
3) CSG and FMF significantly contribute to improving the model's detection performance. Comparing experimental groups Num.1 with Num.3 and Num.4 in Table V reveals enhancements in precision (*P*), recall (*R*), and mean average precision (*mAP*) for medium and small targets, particularly in terms of small target detection accuracy. As discussed in Section IV-A, the combined influence of these two components strengthens the model's information retrieval capability, resulting in the acquisition of high-quality semantic feature maps.

B. Comparison With Mainstream Algorithms

The proposed method's performance is evaluated on the HRSID and SSDD datasets in comparison to several mainstream algorithms. The results are presented in Table VI and the comparative models include faster RCNN+ResNet50 [12], cascade RCNN+ResNet50 [53], SSD300+VGG16, SSD512+VGG16 [54], quad-FPN [39], DAPN [55], ASAFE [56], Pow-FAN [24], and YOLOX [29].

1) Table VI presents the detection results, demonstrating that the proposed model, HRSID and SSDD, outperforms other models with a *P* of 94.81% and 99.01%, and a *mAP* of 92.06% and 99.50% in the HRSID and SSDD datasets, respectively. These results highlight the superior generalization ability of the decoupled-head DCNN algorithm across different datasets.

Synthetic performance (*SP*) is visually compared in Fig. 11, indicating that our proposed model surpasses two-stage target detection algorithms like Cascade RCNN+ResNet50 [53] in accuracy and size. This reduction in training burden is particularly noteworthy. In contrast to anchor frame-based single-stage algorithms, such as ASAFE [56], which can be restrictive to edge delineation accuracy, our anchor frame-free detection strategy effectively enhances accuracy. Our model, with a size of 246.5*M* and a detection latency of 18.1*ms*, is well-suited for real-time monitoring requirements. The addition of the CSG module and the FMF model enriches semantic information, making our algorithm more suitable for SAR target detection compared with other anchor-frame-free mechanisms, such as YOLOX [29].

2) Fig. 12 presents the prediction results of ground truth and the decoupled head-DCNN algorithms on various experimental samples from different datasets. In the case of the HRSID dataset, the model demonstrates strong noise resilience, effectively detecting near-shore targets and densely clustered small targets. However, it may occasionally misjudge small targets due to the predominance of iso-targets in the SAR-Ship-Dataset used for training and the limited land scene area. Conversely, the model excels on the SSDD dataset, characterized by its large scale and uncomplicated background, delivering precise ship boundary delineation for densely arranged ships. Further analysis of common false alarm causes is provided in Section V-B.

3) Fig. 13 presents the results of ship detection in SAR images using various models. The control models encompass

TABLE VI
QUANTITATIVE EVALUATION OF HRSID AND SSDD BY DIFFERENT METHODS

Method	HRSID				SSDD				Model size(M)
	P (%)	R (%)	mAP (%)	Latency (ms)	P (%)	R (%)	mAP (%)	Latency (ms)	
Faster RCNN+ResNet50 [12] *	79.84	83.77	81.80	3	85.07	92.86	92.07	4	330.1
Cascade RCNN+ResNet50 [53] *	83.92	84.43	81.89	3	86.77	92.49	91.61	4	552.5
SSD300+VGG16 [54]	79.51	83.00	79.50	40	82.54	91.76	87.06	36	190.0
SSD512+VGG16 [54]	82.61	83.00	82.50	13	90.27	93.41	89.19	23	195.1
Quad-FPN [39]	87.96	87.92	86.12	13	89.52	95.77	95.29	11	-
DAPN [55]	86.90	81.20	88.20	7	85.60	91.40	90.60	12	-
ASAFE [56]	93.46	86.75	85.18	-	88.54	95.94	95.19	-	-
Pow-FAN [24] +	88.50	83.72	89.74	32	94.61	96.52	96.35	31	136.0
YOLOX [29] +	88.81	80.70	87.72	11	98.07	95.94	96.99	11	-
Ours	94.81	85.66	92.06	18.1	99.09	99.51	99.50	18.5	246.5

In Table VI, "*" represents a two-stage target detection algorithm, "+" stands for anchorless box detection algorithms, and the remaining entries correspond to single-stage target detection algorithms.

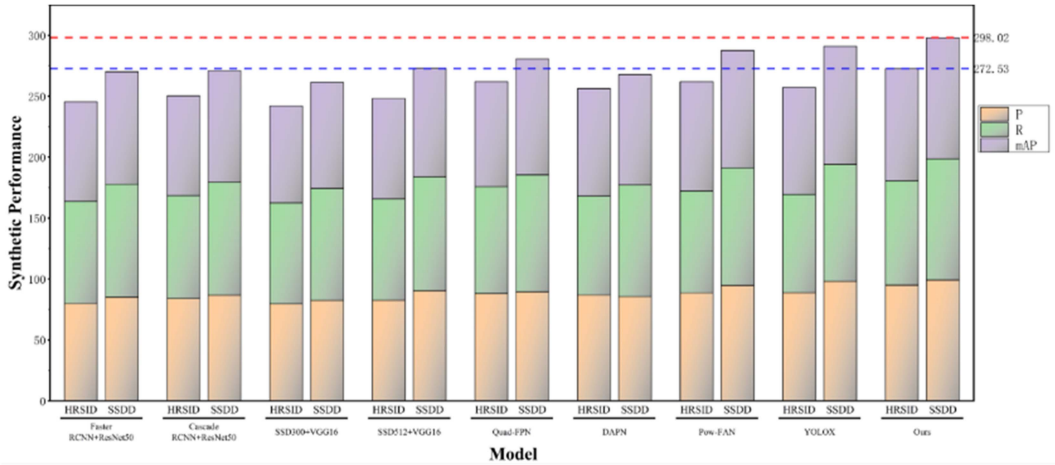

Fig. 11. Comprehensive performance comparison of different detection algorithms on HRSID and SSDD using various detection metrics. (The blue dashed line and red dotted line serve as reference lines for the proposed methods in this article).

two-stage algorithms, single-stage algorithms, and anchorless frame detection-based algorithms. On the SSDD dataset, the control algorithm tends to concentrate on local features with limited global feature learning capability. Similarly, the results from the HRSID dataset show that the control algorithm struggles to accurately detect targets with significant scale differences. In summary, the proposed algorithm improves global localization and boundary feature extraction by incorporating the CSG and FMF modules. Furthermore, it enhances the model's ability to pinpoint the center of ships by employing a decoupled detection head with angle-WIoU correction. These enhancements provide a competitive advantage in terms of detection accuracy compared with other algorithms.

4) Fig. 14 illustrates the precision-recall (PR) curves for two-stage, single-stage algorithms, and the anchorless frame detection algorithm. Our proposed algorithm demonstrates superior detection performance compared with traditional two-stage and single-stage algorithms. Notably, it outperforms the mainstream algorithms employed in our experiments. While the mainstream algorithm prioritizes fast detection through a streamlined network structure, sacrificing accuracy, our proposed model strikes a balance between detection accuracy and speed, enhancing overall performance.

TABLE VII
DETECTION RESULTS OF THE NEARSHORE SHIP DETECTION NETWORK ON THE SSDD DATASET

Method	P(%)	R(%)	AP_{50}(%)
ESD+TDL [57]	91.40	89.70	94.52
OSCD-Net [58]	92.90	88.11	93.98
S^2LSDNet [59]	94.75	85.71	91.80
BSLM-Faster RCNN [60]	87.25	86.89	91.08
BSLM-SSD [60]	90.81	90.27	94.69
Ours-100	**97.50**	**97.50**	**99.13**

C. Algorithm for Nearshore Ship Detection From SAR Images

The comparison methods selected in this part are SAR nearshore ship detection methods, including ESD+TDL [57], OSCD-Net [58], S^2LSDNet [59], BSLM-faster RCNN [60], and BLSM-SSD [60], since most of the networks are pretrained using SSDD dataset, to ensure the fairness of the experiment. The test dataset is selected from the SSDD dataset for testing, setting epoch to 100 and batch size to 12.

1) The detection results of the comparison methods and the proposed method on the SSDD dataset are listed in Table VII,

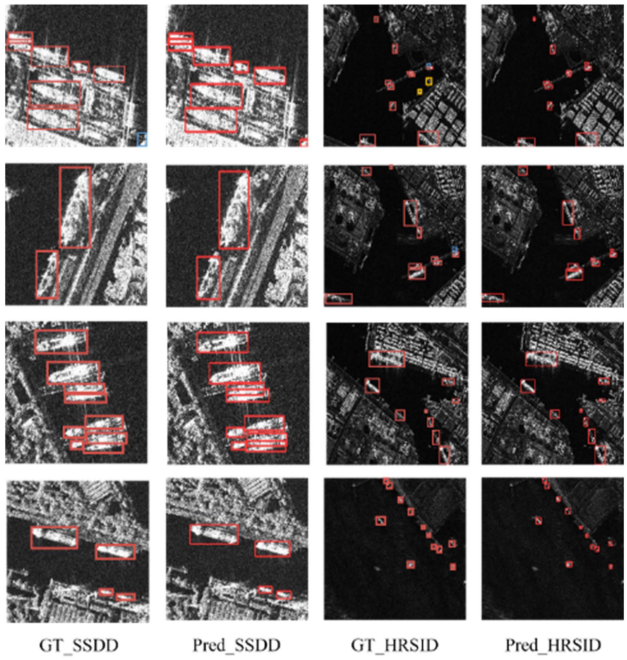

Fig. 12. Detection results of decoupled head-DCNN on SSDD and HRSID. In the *Pred* section, red boxes indicate the detected results, while in the *ground truth* section, red boxes represent sample markers, blue boxes signify instances of misdetection, and yellow boxes indicate instances of missed detection.

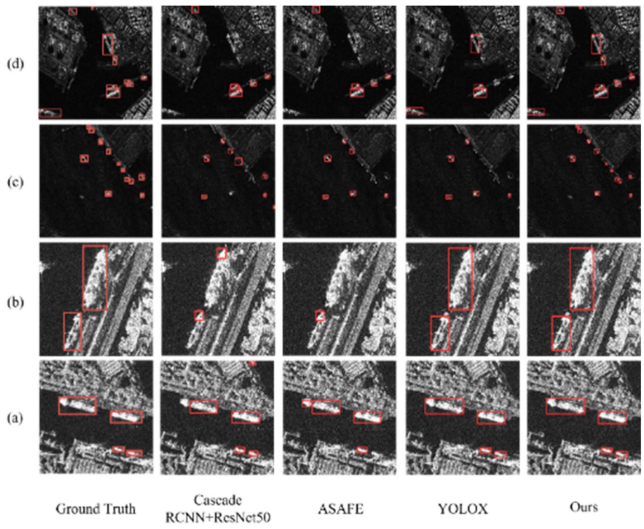

Fig. 13. Detection outcomes of various methods, with columns (a) and (b) representing results from the SSDD dataset and columns (c) and (d) demonstrating results from the HRSID dataset.

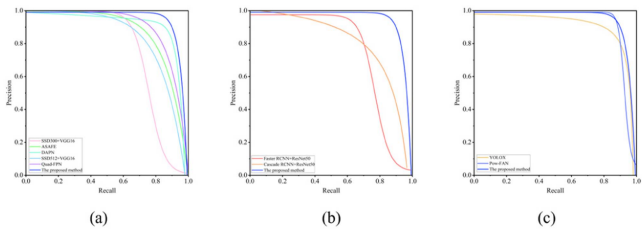

Fig. 14. *PR* curves for diverse methods applied to the HRSID dataset. (a) Selection of two-stage detection algorithms. (b) Various single-stage detection algorithms. (c) Anchorless frame detection algorithms.

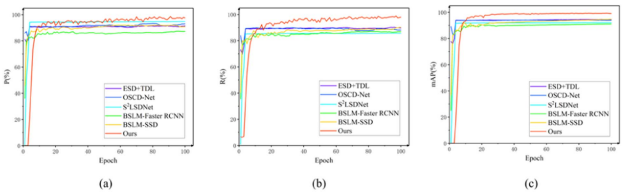

Fig. 15. *P*, *R*, and *mAP* graphs for different algorithms on SSDD.

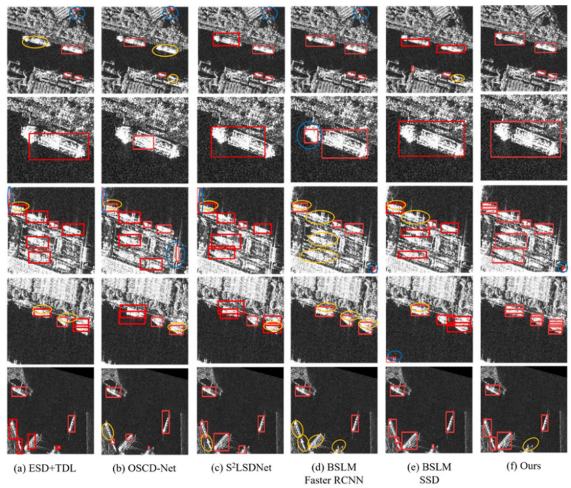

Fig. 16. Detection results of various nearshore SAR ship target detection algorithms.

and the corresponding *P*, *R*, *mAP*, and curves are shown in Fig. 15. Where (a) describes the model precision, (b) describes the model recall, and (c) describes the model average precision, for a more intuitive comparison of the performance of each method. From the table, it can be seen that the proposed method is 2.75% and 7.23% higher than the optimal method in terms of *P* and *R*, respectively. It proves that the proposed method has a significant advantage in SAR near-shore ship detection performance, and the substantial increase in recall is mainly due to the CSG module improving the sensitivity of the model to target features, and the FMF module can effectively fuse features to provide more detailed location description and richer semantic information to distinguish between noise and target features. As a result, the model can show good detection performance.

2) Fig. 16 shows the comparison of the detection results, the yellow and blue circles in the image represent the missed and false alarms, respectively. It can be seen that the proposed model produces fewer misses and false alarms in the dense row of small ships near the shore, the false alarms on the edge of the image are significantly reduced, and the proposed model is more accurate in segmenting ship targets than other methods due to the addition of the angle-WIoU loss function to the proposed method. The comparative method improves the existing network with the idea of deepening the network structure or increasing the attention mechanism to focus on the near-shore ship targets; however, it does not notice that the ship features learned by the network may be similar to some noise features, thus, when some noise features (land, waves) appear on the land or at the edge of the image, the

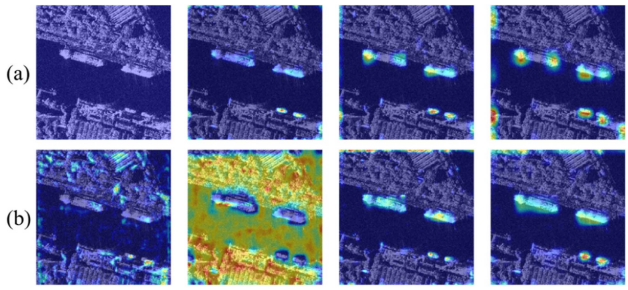

Fig. 17. Feature visualization maps. Feature maps from (a) backbone and (b) Neck.

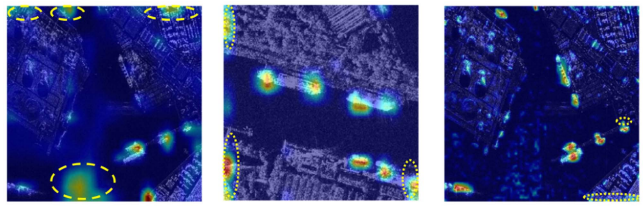

Fig. 18. Decoupled head-DCNN feature map output. The yellow elliptical dashed circles are the attention that may lead to detection errors.

network will be misjudged to be incomplete ship features and thus generate false detections. In addition, certain incomplete ships are not detected, probably because the training samples of incomplete ships are small, and therefore the model is not able to learn such features well.

V. Discussion

A. Feature Visualization

Fig. 17 presents a feature visualization diagram depicting the output from the final C3 module in the backbone. Notably, larger ships exhibit focused attention on the bow and stern, while smaller ships center their attention on the middle section. However, due to the limited semantic information, as observed in the fourth column of the row in Fig. 17(a), certain sea and land features resembling ship features can be misinterpreted, leading to false alarms.

The Neck module effectively fuses the multiscale features extracted from the image, in Fig. 17(b), distinguishing between background environment features and ship features. It separately learns these two sets of features, aiding in the separation of background noise from target features. This approach significantly enhances the network's ability to discern between the two. The results demonstrate that decoupled head-DCNN successfully filters out low-quality semantic information and produces high-quality feature maps through feature extraction and fusion across the backbone and Neck modules.

The feature maps extracted from different convolutional layers may have different receptive fields and represent different semantic information, and usually, the bottom layer feature maps pay more attention to local details, while the top layer feature maps capture more global information, so there will be a difference in the degree of their responses. Second, the proposed method incorporates the CSG attention mechanism in the backbone network, so it will make the difference in the response of the feature maps more obvious, such as the second column of row in Fig. 17(b), which produces feature maps with large differences in the corresponding intensities, and the network's ability for distinguishing between the foreground and the background significantly improves with the continuous learning of these features.

B. Error Analysis

To delve deeper into the origins of false alarms, we perform feature map visualization from the decoupled head-DCNN network using a thermal map, as illustrated in Fig. 18. This analysis focuses on identifying patterns in false alarms and uncovers the primary reasons for these errors, which include the following.

1) In this case, the model appears to learn certain characteristics associated with small and dark ship targets. However, due to the resemblance between some background noise features and these ship target features, this results in misclassification. To address this issue, further investigation can be conducted to differentiate small and dark targets from background noise features, potentially by exploring the concept of pseudo-target detection. Another approach could involve modifying the dataset to label small and dark targets as distinct categories and adjusting weight parameters during dataset creation to enhance network focus.
2) In addition, the image edges are prone to over-attention, which may be due to the fact that the model may mistake the land features as unlearned ship targets during training due to insufficient land background and cause false alarms. In this regard, the inclusion of small sample learning mechanism and unsupervised learning mechanism can be considered to solve the sample insufficiency problem.

VI. Conclusion

SAR-based ship detection has extensive applications in maritime management. To enhance the precision and efficiency of ship detection in SAR images and minimize detection errors, we propose the decoupled head-DCNN algorithm. This algorithm improves the detection of near-shore maritime ship targets by augmenting the network's expressive capabilities through the integration of CSG, FMF, decoupled head, and angle-WIoU loss correction.

To demonstrate the effectiveness and generalizability of the model, we evaluated it using weights trained on the SAR-Ship-Dataset and tested it on the HRSID and SSDD datasets. Our results show that the proposed model performs significantly better compared with the mainstream algorithms, achieving P of 94.81% and 99.01% on the HRSID and SSDD datasets, which are 1.35% and 0.94% higher than the control algorithms, respectively. Similarly, mAP reaches 92.06% and 99.50%, which are

2.32% and 2.51% higher than the control algorithm, respectively. The performance on the SSDD dataset is significantly improved and achieves optimal precision and recall compared with the latest SAR image nearshore vessel detection algorithms.

These experiments showcase the algorithm's capacity to effectively extract multidimensional target features, detect ships in diverse environments, and reduce false positives and omissions near the shore. For further improvement in detection accuracy, we can explore methods like pseudo-target detection for extracting global camouflage features. Additionally, the motion state information of ships may be present in wave features within SAR images, allowing us to investigate how sea background information can be used to infer the ship's movement direction.

REFERENCES

[1] X. Leng, K. Ji, S. Zhou, X. Xing, and H. Zou, "Discriminating ship from radio frequency interference based on noncircularity and non-Gaussianity in Sentinel-1 SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 352–363, Jan. 2019, doi: 10.1109/tgrs.2018.2854661.

[2] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2019, doi: 10.1109/lgrs.2018.2882551.

[3] X. Wang, C. Chen, Z. Pan, and Z. Pan, "Fast and automatic ship detection for SAR imagery based on multiscale contrast measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1834–1838, Dec. 2019, doi: 10.1109/lgrs.2019.2913873.

[4] H. Lin, Z. Shi, and Z. Zou, "Fully convolutional network with task partitioning for inshore ship detection in optical remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1665–1669, Oct. 2017, doi: 10.1109/lgrs.2017.2727515.

[5] A. A. Pappas and D. Bull, "Superpixel-level CFAR detectors for ship detection in SAR imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1397–1401, Sep. 2018, doi: 10.1109/lgrs.2018.2838263.

[6] S. Wang, M. Wang, S. Yang, and L. Jiao, "New hierarchical saliency filtering for fast ship detection in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 351–362, Jan. 2017, doi: 10.1109/tgrs.2016.2606481.

[7] Y. Zhao, L. Zhao, B. Xiong, and G. Kuang, "Attention receptive pyramid network for ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2738–2756, May 2020, doi: 10.1109/jstars.2020.2997081.

[8] X. Leng, K. Ji, X. Xing, S. Zhou, and H. Zou, "Area ratio invariant feature group for ship detection in SAR imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2376–2388, Jul. 2018, doi: 10.1109/jstars.2018.2820078.

[9] H. Lin, S. Song, and J. Yang, "Ship classification based on MSHOG feature and task-driven dictionary learning with structured incoherent constraints in SAR images," *Remote Sens.*, vol. 10, no. 2, p. 190, Jan. 2018, doi: 10.3390/rs10020190.

[10] X. Pan, G. Liao, Z. Yang, and H. Dang, "Ship detection using online update of clutter map based on fuzzy statistics and spatial property," *IET Radar, Sonar, Navigation*, vol. 12, no. 5, pp. 499–505, May 2018, doi: 10.1049/iet-rsn.2017.0471.

[11] X. Leng, K. Ji, S. Zhou, and X. Xing, "Ship detection based on complex signal kurtosis in single-channel SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6447–6461, Sep. 2019, doi: 10.1109/tgrs.2019.2906054.

[12] 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, doi: 10.1109/tpami.2016.2577031.

[13] 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.*, 2016, pp. 779–788, doi: 10.1109/cvpr.2016.91.

[14] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2999–3007, doi: 10.1109/iccv.2017.324.

[15] H. Zhang, Y. Wang, F. Dayoub, and N. Sunderhauf, "VarifocalNet: An IoU-aware dense object detector," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 8510–8519, doi: 10.1109/cvpr46437.2021.00841.

[16] X. Tian, W. Chen, and J. Hu, "Game-theoretic modeling of power supply chain coordination under demand variation in China: A case study of Guangdong Province," *Energy*, vol. 262, Jan. 2023, Art. no. 125440, doi: 10.1016/j.energy.2022.125440.

[17] W. Chen, Z. Shao, K. Wakil, N. Aljojo, S. Samad, and A. Rezvani, "An efficient day-ahead cost-based generation scheduling of a multi-supply microgrid using a modified krill herd algorithm," *J. Cleaner Prod.*, vol. 272, Nov. 2020, Art. no. 122364, doi: 10.1016/j.jclepro.2020.122364.

[18] W. Chen, C. Li, and H. Guo, "A lightweight face-assisted object detection model for welding helmet use," *Expert Syst. Appl.*, vol. 221, no. 1, 2023, Art. no. 119764.

[19] W. Chen, Z. Jiang, H. Guo, and X. Ni, "Fall detection based on key points of human-skeleton using OpenPose," *Symmetry*, vol. 12, no. 5, p. 744, May 2020, doi: 10.3390/sym12050744.

[20] Z. Deng, H. Sun, S. Zhou, and J. Zhao, "Learning deep ship detector in SAR images from scratch," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 4021–4039, Jun. 2019, doi: 10.1109/tgrs.2018.2889353.

[21] Y. Li, W. Zhu, and B. Zhu, "SAR image nearshore ship target detection in complex environment," in *Proc. IEEE 5th Adv. Inf. Technol., Electron., Automat. Control Conf.*, 2021, pp. 1964–1968, doi: 10.1109/IAEAC50856.2021.9391019.

[22] 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, Oct. 2021, Art. no. 4209, doi: 10.3390/rs13214209.

[23] M. F. Humayun, F. A. Nasir, F. A. Bhatti, and M. Tahir, "YOLO-OSD: Optimized ship detection and localization in multiresolution SAR satellite images using a hybrid data-model centric approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 5345–5363, Feb. 2024, doi: 10.1109/JSTARS.2024.3365807.

[24] M. Xiao, Z. He, X. Li, and A. Lou, "Power transformations and feature alignment guided network for SAR ship detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jun. 2022, Art. no. 4509405, doi: 10.1109/lgrs.2022.3183832.

[25] 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, doi: 10.1016/j.patcog.2020.107787.

[26] Q. An, Z. Pan, H. You, and Y. Hu, "Transitive transfer learning-based anchor free rotatable detector for SAR target detection with few samples," *IEEE Access*, vol. 9, pp. 24011–24025, 2021, doi: 10.1109/access.2021.3056663.

[27] X. Wang, Z. Cui, Z. Cao, and S. Dang, "Dense docked ship detection via spatial group-wise enhance attention in SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 1244–1247, doi: 10.1109/igarss39084.2020.9324162.

[28] M. Zhu, G. Hu, S. Li, H. Zhou, S. Wang, and Z. Feng, "A novel anchor-free method based on FCOS + ATSS for ship detection in SAR images," *Remote Sens.*, vol. 14, no. 9, Apr. 2022, Art. no. 2034, doi: 10.3390/rs14092034.

[29] H. Peng and X. Tan, "Improved YOLOX's anchor-free SAR image ship target detection," *IEEE Access*, vol. 10, pp. 70001–70015, 2022, doi: 10.1109/access.2022.3188387.

[30] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," *Int. J. Comput. Vis.*, vol. 128, pp. 642–656, Mar. 2020, doi: 10.1007/s11263-019-01204-1.

[31] Z. Sun et al., "An anchor-free detection method for ship targets in high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7799–7816, Jan. 2021, doi: 10.1109/jstars.2021.3099483.

[32] L. Du, Z. Wang, and Y. Guo, "Adaptive region proposal selection for SAR target detection using reinforcement learning," *J. Radars*, vol. 11, no. 5, pp. 884–896, Nov. 2022, doi: 10.12000/JR22121.

[33] S. Hou, X. Ma, X. Wang, Z. Fu, J. Wang, and H. Wang, "SAR image ship detection based on scene interpretation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 2863–2688, doi: 10.1109/igarss39084.2020.9323473.

[34] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6848–6856, doi: 10.1109/cvpr.2018.00716.

[35] M. A. Rahman and Y. Wang, "Optimizing intersection-over-union in deep neural networks for image segmentation," in *Proc. Adv. Vis. Comput.*, 2016, pp. 234–244, doi: 10.1007/978-3-319-50835-1_22.
[36] L. Bai, C. Yao, Z. Ye, D. Xue, X. Lin, and M. Hui, "Feature enhancement pyramid and shallow feature reconstruction network for SAR ship detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1042–1056, Jan. 2023, doi: 10.1109/jstars.2022.3230859.
[37] Z. Chen and X. Gao, "An improved algorithm for ship target detection in SAR images based on faster R-CNN," in *Proc. 9th Int. Conf. Intell. Control Inf. Process.*, 2018, pp. 39–43, doi: 10.1109/ICICIP.2018.8606720.
[38] Y. Feng et al., "A lightweight position-enhanced anchor-free algorithm for SAR ship detection," *Remote Sens.*, vol. 14, no. 8, Apr. 2022, Art. no. 1908, doi: 10.3390/rs14081908.
[39] T. Zhang, X. Zhang, and X. Ke, "Quad-FPN: A novel quad feature pyramid network for SAR ship detection," *Remote Sens.*, vol. 13, no. 14, Jul. 2021, Art. no. 2771, doi: 10.3390/rs13142771.
[40] J. Kim, J. Choi, M. Cheon, and J. Lee, "MAMNet: Multi-path adaptive modulation network for image super-resolution," *Neurocomputing*, vol. 402, pp. 38–49, 2020.
[41] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur.Conf. Comput. Vis.*, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.
[42] Y. Liu, M. Zhang, P. Xu, and Z. Guo, "SAR ship detection using sea-land segmentation-based convolutional neural network," in *Proc. Int. Workshop Remote Sens. Intell. Process.*, 2017, pp. 1–4, doi: 10.1109/rsip.2017.7958806.
[43] Y. Zhou, X. Jiang, Z. Li, and X. Liu, "SAR target classification with limited data via data driven active learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 2475–2478, doi: 10.1109/igarss39084.2020.9324364.
[44] 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 Observ. Remote Sens.*, vol. 17, pp. 4611–4625, Jan. 2024, doi: 10.1109/JSTARS.2024.3357171.
[45] 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, doi: 10.1109/access.2020.3005861.
[46] T. Zhang et al., "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sens.*, vol. 13, no. 18, Sep. 2021, Art. no. 3690, doi: 10.3390/rs13183690.
[47] S. Lei, D. Lu, X. Qiu, and C. Ding, "SRSDD-v1.0: A high-resolution SAR rotation ship detection dataset," *Remote Sens.*, vol. 13, no. 24, Dec. 2021, Art. no. 5104, doi: 10.3390/rs13245104.
[48] Q. Zhang and Y. Yang, "SA-Net: Shuffle attention for deep convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 2235–2239, doi: 10.1109/icassp39728.2021.9414568.
[49] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10778–10787, doi: 10.1109/cvpr42600.2020.01079.
[50] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, p. 765, Mar. 2019, doi: 10.3390/rs11070765.
[51] A. Bochkovskiy, C. Wang, and H. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[52] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," 2017, *arXiv:1710.09412*.
[53] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6154–6162, doi: 10.1109/cvpr.2018.00644.
[54] W. Liu et al., "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.
[55] Z. Cui, Q. Li, Z. Cao, and N. Liu, "Dense attention pyramid networks for multi-scale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8983–8997, Nov. 2019, doi: 10.1109/tgrs.2019.2923988.
[56] H. Shi, Z. Fang, Y. Wang, and L. Chen, "An adaptive sample assignment strategy based on feature enhancement for ship detection in SAR images," *Remote Sens.*, vol. 14, no. 9, May 2022, Art. no. 2238, doi: 10.3390/rs14092238.
[57] Y. Zhou, F. Zhang, Q. Yin, F. Ma, and F. Zhang, "Inshore dense ship detection in SAR images based on edge semantic decoupling and transformer," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 4882–4890, May 2023, doi: 10.1109/JSTARS.2023.3277013.
[58] J. Zhang, R. Huang, Y. Li, and B. Pan, "Oriented ship detection based on intersecting circle and deformable ROI in remote sensing images," *Remote Sens.*, vol. 14, no. 19, pp. 4749–4769, Sep. 2022.
[59] C. Qin, X. Wang, G. Li, and Y. He, "A semi-soft label-guided network with self-distillation for SAR inshore ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5211814, doi: 10.1109/TGRS.2023.3293535.
[60] T. Zhang et al., "Balance scene learning mechanism for offshore and inshore ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4004905, doi: 10.1109/LGRS.2020.3033988.

Hongjian Ye received the B.E. degree in safety engineering from the Shanghai Institute of Technology, Shanghai, China, in 2022. He is currently working toward the master' degree in safety engineering with the China University of Geosciences, Wuhan, China.

His research interests are object detection and computer vision.

Weiming Chen received the Ph.D. degree in mechanical science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2011.

He was a Postdoctoral Fellow with Geological Resources and Geological Engineering, China University of Geosciences, Wuhan. He is currently an Associate Professor and Master Supervisor with the Faculty of Engineering and Institute for Natural Disaster Risk Prevention and Emergency Management, China University of Geosciences. His research interests are intelligent emergency technology and equipment, intelligent perception of security situation, and intelligent safety supervision and service.

Diyong Wang received the B.E. degree in safety engineering from Sichuan Normal University, Chengdu, China, in 2021. She is currently working toward a master's degree in safety science and engineering with the China University of Geosciences, Wuhan, China.

Her research interests are feature extraction and computer vision.

Yu Qiu received the B.E. degree in safety engineering in 2023 from the China University of Geosciences, Wuhan, China, where he is currently working toward the master's degree in safety science and engineering.

His research interests are feature engineering and time course prediction.