# Scale-Mixing Enhancement and Dual Consistency Guidance for End-to-End Semisupervised Ship Detection in SAR Images

Man Chen, Yuanlin He, Tianfeng Wang, Yahao Hu, Jun Chen, and Zhisong Pan

***Abstract*—The object detection of synthetic aperture radar (SAR) ships holds significant promise for water traffic monitoring, ship search and rescue, and maritime warning tasks. Regrettably, most of the existing SAR ship object detection is limited to the fully supervised paradigm, exhibiting a firm reliance on data labels, and inherent challenges such as multiscale ship feature disparities and indistinct small-sized ships also make SAR ship detection difficult. To address these, we propose a scale-mixing enhanced and dual consistency guided semisupervised object detection (SMDC-SSOD) method. Specifically, this method is based on the teacher–student framework and primarily comprises three core components: cross-scale feature mixing (CSFM) scheme, scale change consistency guidance (SCCG) strategy, and proposal consistency guidance (PCG) strategy, which can efficiently conduct end-to-end semisupervised learning from limited data labeling, achieving low-cost and high-performance ship perception. CSFM scheme includes interpyramid and intrapyramid feature cross-scale mixings, which can improve the network's adaptability for multiscale ship characteristics and increase focus on small-sized ships. SCCG strategy leverages variations in confidence scores at different scales to select valuable pseudolabels, providing more precise guidance for the student network. PCG strategy further reflects the positioning quality of pseudolabels through the proposal consistency generated by the student network, guiding it to make high-quality predictions. The experimental results on the publicly available HRSID, BBox-SSDD, and SAR-Ship-Dataset demonstrate that SMDC-SSOD can accurately detect SAR ships with an extremely low data annotation rate (below 10%) and achieve optimal detection performance compared to state-of-the-art methods.**

***Index Terms*—Dual consistency guidance, scale-mixing enhancement, semisupervised object detection (SSOD), synthetic aperture radar (SAR), teacher–student frame (TSF).**

## I. Introduction

SYNTHETIC aperture radar (SAR) is an advanced active microwave sensor that offers unique advantages, such as independence from time, light, and weather conditions, compared to other sensors such as optical, infrared, and hyperspectral [1]. This makes it particularly suitable for monitoring diverse maritime environments. SAR ship detection, an essential component of maritime surveillance tasks, aims to accurately perceive ship information from SAR images, aiding in water traffic monitoring, ship search and rescue, maritime warnings, and port scheduling, and has garnered widespread attention.

In recent years, with the advancement of deep learning (DL) and improved computational capabilities, natural image object detection methods have rapidly evolved, giving rise to a variety of DL-based object detection techniques, including Faster R-CNN [2], Cascade R-CNN [3], EfficientDet [4], and DETR [5]. Benefiting from the development of natural image detection methods, DL-based detection approaches have also been extended to ship detection in SAR images. These methods can utilize DL models to learn feature representations from large-scale SAR ship datasets, thereby achieving the perception of ship targets within SAR images. As research progresses, many DL-based SAR image ship detection methods have also placed significant emphasis on the imaging mechanisms of SAR images and the intrinsic characteristics of ships, further enhancing detection performance [6], [7], [8], [9], [10], [11], [12].

However, current DL-based SAR ship detection methods are predominantly based on fully supervised paradigms, relying on many high-quality labels and imposing high manual annotation costs. Therefore, we shift focus from fully supervised SAR ship object detection toward semisupervised object detection (SSOD), aiming to achieve SAR ship detection using only a small number of labeled annotations and reduce dependence on large-scale data labeling. As seen from the development lineage illustrated in Fig. 1(a), this work represents a more in-depth study compared to natural image detection and fully supervised SAR ship detection tasks.

Compared to conventional object detection, SSOD has relatively weak supervisory information, presenting challenges for accurately perceiving targets. Existing SSOD methods primarily rely on the teacher–student framework (TSF) [13], [14], [15], which can utilize a teacher network to generate pseudolabels for unlabeled data and guide the student network for comprehensive training, promoting precise target perception. Based on differences in training approach, SSOD methods can be divided into multistep training-based SSOD methods [16], [17], [18] and end-to-end training-based SSOD methods [19], [20], [21], [22]. In multistep training-based SSOD methods, the teacher network initially undergoes separate training on labeled data to achieve preliminary learning and subsequently predicts unlabeled data to generate pseudolabels for training the student network.

Received 11 June 2024; revised 30 July 2024; accepted 21 August 2024. Date of publication 29 August 2024; date of current version 13 September 2024. *(Corresponding author: Zhisong Pan.)*

The authors are with the School of Command and Control Engineering, Army Engineering University of PLA, Nanjing 210007, China (e-mail: chenmanhjdx@163.com; cathy345543@163.com; 837706115@qq.com; huyahao@aeu.edu.cn; chenjun_sc@aeu.edu.cn; panzhisong@aeu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3451549

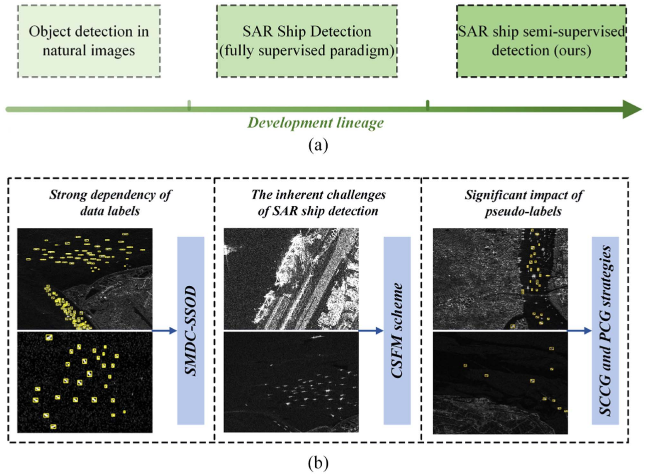

Object detection in natural images
SAR Ship Detection (fully supervised paradigm)
SAR ship semi-supervised detection (ours)
Development lineage
(a)
Strong dependency of data labels
The inherent challenges of SAR ship detection
Significant impact of pseudo-labels
SMDC-SSOD
CSFM scheme
SCCG and PCG strategies
(b)

Fig. 1. (a) Development lineage from natural image object detection to SAR ship detection to semisupervised SAR ship detection. (b) Factors emphasized in the SAR ship detection task and the proposed response approaches in this research.

However, due to the involvement of multiple training steps, the training process for multistep training-based SSOD methods is intricate. Their detection performance may also be limited by the teacher network initialized solely based on labeled data. In contrast, end-to-end training-based SSOD methods optimize both teacher and student networks throughout the entire training process, resulting in a more straightforward training process and easier attainment of effective detection performance, thus garnering broader attention. In this research, we specifically focus on end-to-end training-based SAR ship SSOD methods, aiming to achieve efficient end-to-end semisupervised training using a small amount of data labels, facilitating a high-performance perception of SAR ships at low cost.

Furthermore, whether it is multistep training-based SSOD methods or end-to-end training-based SSOD methods, pseudolabels directly influence the learning quality of the student network. High-quality pseudolabels provide accurate and comprehensive guidance for the student network, promoting precise bounding box predictions. In contrast, a scarcity or poor quality of pseudolabels leads to inadequate or inaccurate supervision information for the student network, affecting detection performance. Therefore, emphasizing pseudolabels is crucial in the design of SAR ship SSOD methods. In addition, SAR ship detection tasks pose inherent challenges compared to natural scene object detection. On the one hand, due to the diversity in the physical sizes of ships and the variability in SAR image resolutions, ships in SAR images often exhibit multiscale feature disparities, posing difficulties in ship detection [23], [24], [25]. On the other hand, due to the low proportion of pixels occupied by small-sized ships in SAR images, they often appear indistinct and are susceptible to interference from clutter and speckle noise, among other distractions [26], [27], [28]. Consequently, considering these inherent challenges in the design process of SSOD methods for SAR ships is appropriate.

Taking into account the strong dependence on data labels, the significant impact of pseudolabels, and inherent challenges such as multiscale ship feature disparities and indistinct small-sized ships, we focus on the SSOD method for SAR ship detection, proposing a scale-mixing enhanced and dual consistency guided semisupervised object detection method, namely SMDC-SSOD. Fig. 1(b) summarizes the key factors considered in the SAR ship detection task and the approach presented in this research. Specifically, this method, based on the TSF, primarily comprises three core components: the cross-scale feature mixing (CSFM) scheme, the scale change consistency guidance (SCCG) strategy, and the proposal consistency guidance (PCG) strategy, which can enable efficient end-to-end semisupervised training using a limited amount of data labels, achieving low-cost, high-performance perception of SAR ships. The CSFM entails interpyramid and intrapyramid feature cross-scale mixings, enhancing the network's adaptability to multiscale ship feature disparities and increasing its focus on small-sized ships. The SCCG strategy utilizes changes in confidence scores across different scales to design pseudolabel filtering criteria, aiding in selecting more valuable pseudolabels from teacher network predictions and providing more precise guidance for the student network. The PCG strategy leverages the consistency of proposals generated by the student network to reflect the localization quality of pseudolabels further, guiding the student network to make high-quality bounding box predictions. The main contributions can be summarized as follows.

1) We propose SMDC-SSOD based on the TSF framework, which enables efficient end-to-end semisupervised learning using a limited amount of data labels, achieving low-cost, high-performance perception of ships in SAR images.
2) We construct the CSFM scheme, encompassing interpyramid and intrapyramid feature cross-scale mixings, to enhance the network's adaptability to multiscale ship feature disparities and increase its focus on small-sized ships.
3) The SCCG and PCG strategies are designed to create pseudolabel filtering criteria and reflect the localization quality of pseudolabels, thus providing more precise guidance for the student network and enabling higher quality bounding box predictions.
4) Results from experiments conducted on the high-resolution SAR images dataset (HRSID), bounding box SAR ship detection dataset (BBox-SSDD), and SAR-Ship-Dataset demonstrate that SMDC-SSOD accurately detects ships in SAR images with annotation rates as low as 10% outperforming state-of-the-art semisupervised detection methods.

## II. Related Work

### A. Object Detection

Mainstream object detection methods include two architectures: convolutional neural networks (CNN) and the transformer. Methods based on the CNN architecture can be further subdivided into two-stage methods [2], [3] and one-stage detection methods [4], [29], [30]. The most representative two-stage detection method is Faster R-CNN [2], which first generates initial regions of interest (RoI) through a region proposal network (RPN) and then performs specific classification and regression predictions using head networks. Building upon Faster R-CNN, Cascade R-CNN [3] introduces a cascading structure, employing a series of subnetworks for multistage perception, further enhancing object detection performance. One-stage methods directly perform dense classification and bounding box regression on input images without explicitly generating candidate regions. EfficientDet [4] achieves improved feature representation by redesigning the feature pyramid network (FPN) [31] and proposes a compound scaling method that unifies scale, depth, and width simultaneously, efficiently utilizing computational resources. QGL-G [30] is a one-stage object detection method that adapts to imbalanced positive and negative samples. The gradient-enhanced function and quality-guided loss in this method strengthen the utilization of positive samples, ultimately achieving high efficiency and accuracy.

In recent years, object detection methods based on the transformer architecture have also garnered attention. Specifically, DETR [5] regards the object detection task as a set prediction problem, incorporating a set-based global loss to enforce one-to-one matching and enabling direct parallel output of the final prediction set. Based on DETR, H-DETR [32] adopts a mixed matching scheme, combining original one-to-one matching with auxiliary one-to-many matching during training to improve detection accuracy. The aforementioned methods of both architectures are designed for natural image applications. Given that SAR images have unique imaging mechanisms and the characteristics of ships differ from objects in natural images, applying these methods to SAR ship detection tasks may be subject to certain limitations.

### B. Ship Detection in SAR Images

Early-stage SAR ship detection was primarily achieved through handcrafted features, including the CFAR method [33], global thresholding method [34], polarization decomposition method [35], and visual saliency detection method [36]. The CFAR method [33] is widely used, mainly determining an adaptive threshold with a constant false alarm probability based on statistical characteristics of sea clutter, then employing a sliding window to search for ships within the background window. The global thresholding method [34] first establishes a global threshold using statistical decision-making methods and then searches bright targets across the entire SAR image to extract RoI. The polarization decomposition method [35] detects ships in polarimetric SAR images by exploiting the differences in the backscattering characteristics between ship targets and sea clutter. As for the visual saliency detection method [36], it initially extracts salient regions from the entire scene and then utilizes local feature relationships in space to construct the visual saliency map further, thereby uncovering ships. The handcrafted feature methods lack scene generalization and require complex theoretical support and extensive human involvement.

Benefiting from the advancements in object detection methods for natural images, DL-based SAR ship detection methods have also made significant strides. These methods are capable of leveraging large-scale SAR ship datasets to learn high-performance feature representations, taking into account the imaging mechanisms of SAR images and the characteristics of ships [6], [7], [8], [9], [10], [11], [12]. Specifically, Zhang et al. [6], addressing scenes with clutter interference in SAR images, proposed a frequency attention mechanism to adaptively process frequency-domain information in SAR images adaptively, thus suppressing sea clutter and enhancing adaptability to challenging scenarios. Sun et al. [7] integrated attention feature fusion, depthwise separable convolution, interlayer connections, and multiscale strategies into the network structure, enabling the network to exhibit high precision and efficiency while demonstrating outstanding performance in detecting small targets within SAR images. Huang et al. [8], addressing the inability of existing SAR ship detectors to express reliability and interpretability, constructed a Bayesian deep detector to quantify uncertainty and designed an occlusion-based explanation method to explain SAR scattering features, promoting the development of interpretable and trustworthy models in the SAR domain. Wan et al. [9] achieved efficient directional detection of ships in SAR images by improving model feature extraction, boundary perception, and label matching. Zhou et al. [10] introduced a novel SAR ship detection model called the balanced feature enhanced attention model, which incorporates the advantages of attention mechanisms and multiscale feature fusion, demonstrating good adaptability to complex background interference and varying ship sizes within SAR images. Building upon a one-stage object detector, Li et al. [11] proposed a new ship detection network in SAR images. This network robustly extracts features using a multilevel pyramid and enhances the network's adaptability to ship objects by utilizing convolutional channel attention and task decoupling operations. Overall, the aforementioned SAR ship detection methods are mainly within the fully supervised paradigm, exhibiting a strong dependence on data labels, which undoubtedly requires a substantial amount of manual annotation costs. This research shifts attention from fully supervised SAR ship object detection to SSOD, aiming to achieve SAR ship detection tasks using a limited number of data labels, thereby reducing label production costs.

### C. Semisupervised Object Detection

Driven by the success of semisupervised classification, SSOD has also received preliminary research attention in recent years. Compared to conventional object detection, the most significant feature of SSOD is its relatively weak supervisory information, which presents challenges for accurate target perception. Therefore, existing SSOD methods primarily rely on the TSF [13], [14], [15] for construction. TSF typically consists of teacher and student networks, sharing similar structures but with different weights. During semisupervised training, the teacher network first generates pseudolabels for unlabeled data. These pseudolabels are then used as ground truth to supervise the student network, facilitating comprehensive training and achieving accurate target perception.

Existing SSOD methods can be categorized into two types based on their training approaches: multistep training-based

SSOD [16], [17], [18] and end-to-end training-based SSOD [19], [20], [21], [22]. Earlier SSOD methods primarily relied on multistep training, where the teacher network is initially trained separately on labeled data to achieve initial learning, followed by predicting unlabeled data to generate pseudolabels for training the student network. Among these, STAC [16] enriched pseudolabel information through data augmentation to facilitate comprehensive knowledge learning by the student network from unlabeled images. Unbiased Teacher [17] introduced exponential moving average (EMA) [37] into the training process to enhance training stability. ASTOD [18] supplemented an additional refining learning step, further training the student network with labeled data after pseudolabel training to enhance learning effectiveness. The training steps of those above multistep training-based SSOD are cumbersome, and the teacher network is initialized solely based on labeled data, which may limit detection performance.

In contrast to multistep training-based SSOD, end-to-end SSOD can optimize both the teacher and student networks throughout the training process, resulting in better learning effectiveness. Soft Teacher [19] integrated the reliability measure of pseudolabels generated by the teacher network as classification scores into the design process of the loss function during end-to-end training, promoting emphasis on high-quality pseudolabels. PseCo [20] learned more robust feature representations from unlabeled data through feature-level consistency training and improved the pseudolabel filtering strategy by exploring more profound rules, thus enhancing detection performance. LabelMatch [21] guided pseudolabel generation by computing the label distribution of labeled data and further mined potential valuable pseudolabels from dense proposals through label mining to guide the student model in making more precise bounding box predictions. A few recent works have attempted to introduce SSOD into the domain of SAR ship perception. Specifically, Zhou et al. [38] proposed the first end-to-end semisupervised SAR ship detection method, which enhances model robustness through an interference consistency learning mechanism and incorporates background knowledge surrounding SAR ships to calibrate pseudolabels, achieving effective detection of ships within SAR images. Tian et al. [39] designed a hard-sigmoid function to weigh the loss of pseudolabeled data and mitigate the negative impact of pseudolabels on the training process and further integrated high-quality pseudolabels into retraining by using a subnetwork for awarding intersection over union (IoU), thereby promoting the enhanced perception of objects within SAR images. In this study, we further delve into the research of SSOD methods and consider inherent challenges such as multiscale ship feature disparities and indistinct small-sized ships, aiming to achieve cost-effective and high-quality perception of SAR ships.

## III. Methodology

### A. Problem Description

This research introduces semisupervised learning into SAR ship object detection, aiming to train the network with a small amount of labeled data and a large amount of unlabeled data to detect ships in images. The training dataset $D$ comprises two parts: labeled data $D_l = \{I_l^i, L_l^i\}_{i=1}^{N_l}$ and unlabeled data $D_u = \{I_u^i\}_{i=1}^{N_u}$, where $N_l$ and $N_u$ represent the image quantity in labeled and unlabeled data, respectively. Typically, the image quantity $N_u$ in unlabeled data is significantly greater than the image quantity $N_l$ in labeled data, and it is assumed that labeled data $D_l$ and unlabeled data $D_u$ share the same distribution. For an image $I_l^i$ in the labeled data, its corresponding annotation $L_l^i$ consists of bounding box labels and class labels for the objects in the image. For an image $I_u^i$ in the unlabeled data, pseudolabels denoted as $L_p^i$ are generated by the teacher network for training the student network. In summary, SAR ship SSOD trains the model by jointly using a small amount of labeled data $D_l$ and a large amount of unlabeled data $D_u$ to achieve cost-effective and high-performance SAR ship perception.

### B. Overview of the Proposed SMDC-SSOD

The proposed SMDC-SSOD structure follows TSF, consisting of two detectors: a teacher and student networks. The training of SMDC-SSOD involves supervised and unsupervised learning. Fig. 2(a) and (b) represents the overall framework of these two parts. In the supervised learning part, the student network is trained with the labeled data $D_l$, following the same training approach as conventional fully supervised detectors. The unsupervised learning part first subjects the unlabeled data $D_u$ for weak and strong data augmentation to obtain enhanced data $D'_u$ and $D''_u$, respectively, to augment the asymmetry between the teacher and student networks and construct effective training signals. Subsequently, the weak enhanced data $D'_u$ and strong enhanced data $D''_u$ are individually input to the teacher and student networks. The teacher network generates pseudolabels for the unlabeled images, guiding the student network through these pseudolabels to achieve high-quality ship detection. During the training phase, the teacher network updates model weights from the student network via EMA [37], promoting training stability. In the training process of SMDC-SSOD, supervised and unsupervised learning co-occur, randomly sampling labeled and unlabeled images at a preset ratio to form a data batch, enabling end-to-end training.

The core components of SMDC-SSOD include the CSFM scheme, SCCG strategy, and PCG strategy. Among these, the CSFM scheme acts on both the teacher and student networks, comprising interpyramid feature cross-scale mixing and intrapyramid feature cross-scale mixing. The former enhances the network's adaptability to multiscale ship feature differences by blending pyramid features from different views. At the same time, the latter efficiently captures significant semantic information using content-aware feature reassembly and passes it level by level to the bottom of the pyramid, achieving a thorough mixing of intrapyramid features and increasing attention to small-sized ship features. The SCCG strategy provides a pseudolabel filtering indicator based on variations in confidence scores at different scales to help the teacher network filter valuable pseudolabels from predictions of the teacher network, thereby providing more precise guidance to the student network. The PCG strategy further reflects the quality of pseudolabels by

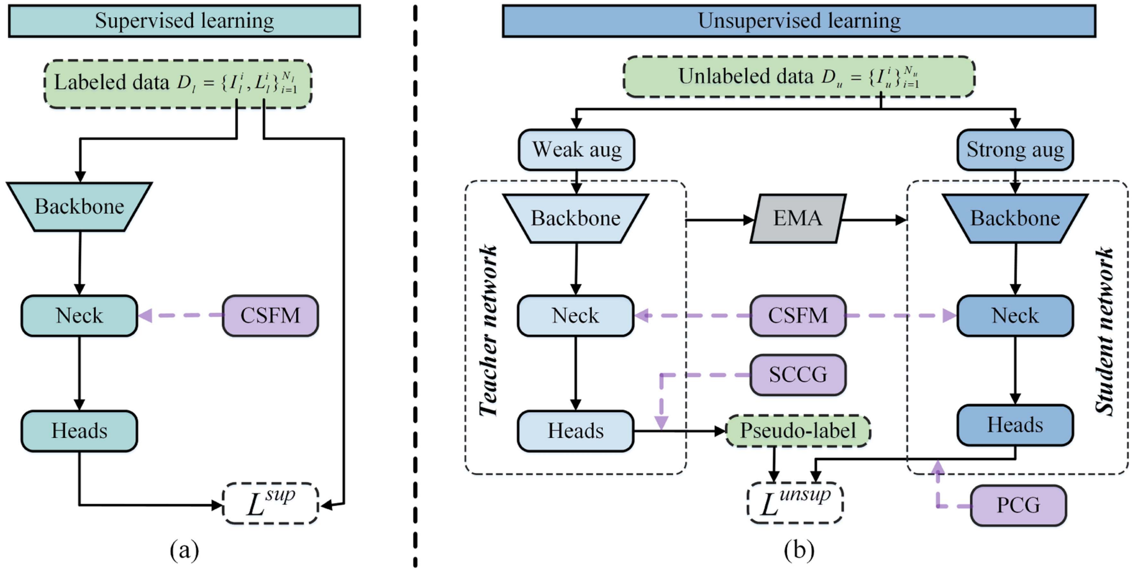


Fig. 2. Overall framework of SMDC-SSOD. (a) Overall framework of the supervised learning part. (b) Overall framework of the unsupervised learning part. For brevity, we did not depict the RPN part within the network.

leveraging the consistency of proposals generated by the student network, guiding the student network to make high-quality bounding box predictions.

### *C. Cross-Scale Feature Mixing Scheme*

Compared with object detection in natural scenes, SAR ship detection tasks often exhibit multiscale feature differences due to the diversity of ship physical sizes and SAR image resolutions. Furthermore, small ships in SAR images often appear indistinct due to their limited pixel occupancy, quickly submerging them in sea clutter and speckle noise. To address this, we propose the CSFM scheme to achieve CSFM from interpyramid and intrapyramid perspectives. The interpyramid feature cross-scale mixing constructs richer feature representations with the help of downsampled views and achieves efficient feature interactions by mixing feature pyramids in different views, improving the network's adaptability to the differences in multiscale ship features. Intrapyramid CSFM utilizes content-aware feature reassembly to fully capture significant semantic information, progressively propagating semantic information down to the pyramid base to achieve comprehensive feature mixing, enhancing focus on small-sized ship features.

In terms of interpyramid CSFM, akin to the idea of enhancing perceptual capabilities by incorporating additional views [22], [40], we first downsample the input network's image $I$ by a factor of 0.5 to obtain its downsampled view represented as $I_d$. If the input network is the teacher network, then $I \in D'_u$; if the input network is the student network, then $I \in D''_u$. Subsequently, $I$ and $I_d$ are subjected to feature extraction separately to obtain conventional-scale feature pyramid $P_c = \{p_c^i\}_{i=2}^6$ and small-scale feature pyramid $P_s = \{p_s^i\}_{i=2}^6$. Next, dynamic weighted fusion is applied to adjacent-level features $P_c$ and $P_s$ at the same scale in conventional-scale feature pyramid $p_c^i$ and small-scale feature pyramid $p_s^i$ to achieve cross-scale feature fusion, yielding the mixed-scale feature pyramid $P_m = \{p_m^i\}_{i=2}^6$. It is worth noting that since feature $p_c^2$ at the conventional-scale feature pyramid $P_c$ does not exist at the adjacent level at the same scale in the small-scale feature pyramid $P_s$, feature $p_m^2$ in the mixed-scale feature pyramid $P_m$ is a direct copy of $p_c^2$ from the conventional-scale feature pyramid $P_c$. While training on unlabeled data, the teacher network generates higher quality pseudolabel $L_p$ through the cross-scale mixed feature pyramid $P_m$ because it contains rich multiscale information; the student network inputs the small-scale, mixed-scale, and conventional-scale feature pyramids $P_s$, $P_m$, and $P_c$, respectively, into the head network to obtain three-scale outputs $R_s$, $R_m$, and $R_c$. Ultimately, by treating the pseudolabel $L_p$ generated by the teacher network as ground truth, we efficiently supervise the student network's output predictions $R_s$, $R_m$, and $R_c$, achieving multiscale training. Fig. 3 illustrates the interpyramid feature cross-scale mixing in the CSFM scheme. The small-scale and mixed-scale feature pyramids are only used during training to assist the network in efficient perception. As shown by the red curve in Fig. 3, during the testing phase, the student network removes the small-scale and mixed-scale components, only employing the conventional-scale feature pyramid $P_c$ for prediction and considering the prediction result $R_c$ corresponding to $P_c$ as the final output to reduce model complexity.

In terms of intrapyramid CSFM, we efficiently capture salient semantic information in the feature map over a sizeable receptive field through content-aware feature reassembly [41] and progressively propagate the semantic information down to the pyramid base, achieving comprehensive feature fusion within the pyramid and enhancing focus on small-sized ship features. As shown in Fig. 4, content-aware feature reassembly involves two steps: kernel prediction and feature reassembly. Taking the spatial position $l = (i, j)$ in the input feature map $F$ as

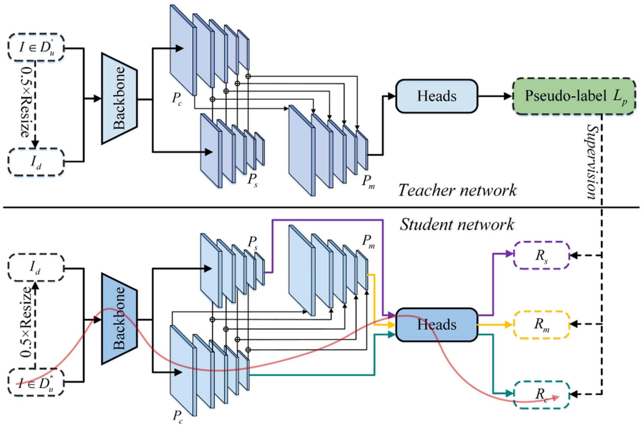


Fig. 3. Illustration of the interpyramid feature cross-scale mixing in the CSFM scheme. The red curve represents the inference process during the testing phase.

t o t a l

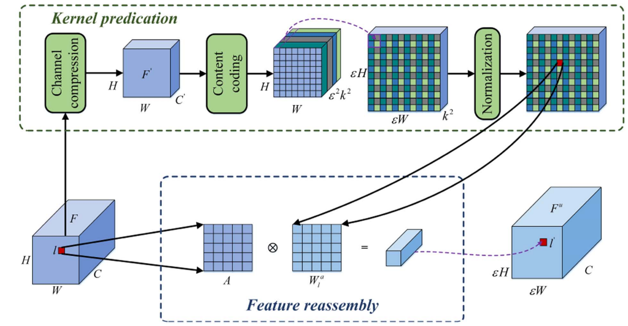


Fig. 4. Schematic diagram of the content-aware feature reassembly in intrapyramid feature cross-scale mixing.

an example, the kernel prediction process aims to generate an adaptive feature reassembly kernel $W_l$ for the spatial position $l$ in a content-aware manner based on the $k \times k$ neighborhood of the feature $F_l$ at the spatial position $l$ in the input feature map $F$. Specifically, for an input feature map $F$ with dimensions of $H \times W \times C$, a $1 \times 1$ convolution is first used to compress its channel dimension, resulting in a feature map $F'$ with dimensions of $H \times W \times C'$, reducing parameters and computational load. Subsequently, the compressed feature map $F'$ is encoded using a content encoding operation to obtain a reassembly kernel with dimensions of $H \times W \times (\varepsilon^2 k^2)$, where $\varepsilon$ represents the upsample ratio. The content encoding operation here is constructed through a convolution with a kernel size of $k$, which can effectively leverage the content information of input features and achieve a richer feature representation. Next, the pixel shuffling operation [42] reshapes the reassembly kernel to $\varepsilon H \times \varepsilon W \times k^2$. Finally, the softmax function is applied to normalize the reassembly kernel, obtaining the weight of each subcontent relative to the total content and acquiring the adaptive feature reassembly kernel $W^a$. Therefore, for the feature $F_l$ at the spatial position $l$ in the input feature map $F$, the aforementioned kernel prediction process can be represented by the following formula:

$$W_l^a = \mathrm{softmax}(\mathrm{shffle}(f_e(f_c(F_l|\theta_c)|\theta_e))) \tag{1}$$

where $W_l^a$ represents the reassembly kernel weights corresponding to $F_l$, $f_e(\cdot|\theta_e)$ is the content encoding operation, and $f_c(\cdot|\theta_c)$ represents the channel compression operation.

Furthermore, the feature reassembly step utilizes the reassembly kernel to generate the final upsampled feature map $F^u$ from the input feature map $F$, as depicted in Fig. 4. Taking the spatial position $l = (i, j)$ in the input feature map $F$ as an example, the calculation of the output $F_{l'}^u$ after feature reassembly is as follows:

$$F_{l'}^u = \sum_{n \in A} \sum_{m \in A} W_{l,(n,m)}^a F_{(i+n,i+m)} \tag{2}$$

where $l'$ represents the spatial position in the upsampled feature map $F^u$ corresponding to the spatial position $l$ in the input feature map $F$, and $A$ denotes the $k \times k$ neighborhood around the feature $F_l$. By predicting feature upsampling output using content-aware adaptive reassembly kernel $W^a$, we can efficiently focus more on significant semantic information near the original spatial position over a larger receptive field, thus obtaining a more vital upsampled feature map $F^u$ than the original feature map. We utilize those above high-quality upsampled feature map $F^u$ to replace the nearest-neighbor interpolated output in both the conventional-scale and small-scale feature pyramids of the teacher and student networks, facilitating efficient propagation of strong semantic information from the top to the bottom of the pyramid, thus achieving thorough attention to small-sized ship features.

### *D. Scale Change Consistency Guidance Policy*

In SSOD, the pseudolabels generated by the teacher network directly impact the learning quality of the student network. High-quality pseudolabels can effectively guide the student network, improving its performance in boundary box prediction. The insufficient quantity or poor quality of pseudolabels can result in inadequate or inaccurate supervision for the student network, consequently affecting the detection performance. Previous work primarily filtered the predictions of the teacher network using a fixed high-confidence score threshold to generate pseudolabels for the teacher network to learn. This approach is overly simplistic and may lead to valuable pseudolabels being overlooked. Inspired by [22], we introduce the SCCG strategy, which designs pseudolabel filtering criteria based on the consistency of confidence scores at different scales, aiding SMDC-SSOD in selecting more valuable pseudolabels from the predictions of the teacher network, thus providing more precise guidance for the student network.

Specifically, existing object detectors often enhance their ability to describe objects by assigning multiple proposals. Therefore, we initially treat each low-quality candidate output from the RPN in the teacher network as a bag of proposals, denoted as $Q = \{q_i\}_{i=1}^{N_q}$, where $N_q$ represents the number of proposals in the bag of predictions. Subsequently, taking the RoI corresponding to proposal $q_i$ and the FPN as inputs to the classification head in the teacher network, we obtain the confidence scores $S_r^i$ and $S_m^i$ for the proposal $q_i$ at standard scale and mixed scale, expressed as follows:

$$S_r^i = h_c(R(q_i), P_r|\theta_c) \tag{3}$$

$$S_m^i = h_c(R(q_i), P_m|\theta_c) \tag{4}$$

where $h_c(\cdot|\theta_{h_c})$ denotes the classification head within the teacher network, $R(q_i)$ represents the RoI corresponding to the proposal $q_i$, and $P_c$ and $P_m$ represent FPN at conventional and mixed scales, respectively. Finally, we calculate the average confidence score improvement for each proposal from conventional scale to mixed scale, considering it as the filtering criterion for pseudolabels, which is computed as follows:

$$\Delta S = \frac{1}{N_q}\sum_{i=1}^{N_q}(S_m^i - S_r^i). \tag{5}$$

When $\Delta S$ is relatively high, indicating that the scale mixing operation can bring substantial benefits, we consider the prediction result valuable for enhancing the student network's adaptability to scale variations. Thus, prediction results with $\Delta S$ more significant than the threshold $\varphi$ is included in the pseudolabels, providing guidance for the student network and high-confidence prediction results, thereby facilitating better prediction outcomes. In general, SCCG benefits from the cross-scale feature fusion across pyramids in the CSFM scheme, and by leveraging changes in confidence scores at different scales, it designs filtering criteria for pseudolabels, enabling further exploration of more valuable pseudolabels from the teacher network predictions and providing more prosperous guidance for the student network.

### E. Proposal Consistency Guidance Policy

In the unsupervised learning component of SMDC-SSOD, the high-score threshold and SCCG strategy can effectively extract high-quality pseudolabels from the teacher network's predictions to provide supervisory information for the student network. Considering that the consistency between pseudoboxes and the corresponding proposals generated by the student network can, to some extent, reflect the quality of the pseudoboxes [40], we design the PCG strategy to guide the student network through the consistency of proposals generated by the student network, aiming to facilitate high-quality boundary box predictions and further enhance its detection performance.

Specifically, if a pseudolabel box exhibits high consistency with a series of proposals generated by the student network, it indicates higher quality. Taking pseudobox $b$ as an example, its consistency $C^b$ with proposals generated by the student network can be calculated by the following formula:

$$C^b = N_p^{-1} \cdot \sum_{i=1}^{N_p} g_i^b \tag{6}$$

where $g_i^b$ represents the consistency between pseudobox $b$ and the $i$th proposal generated by the student network, and $N_p$ denotes the number of positive samples assigned to pseudobox $b$, which acts as a normalizer. In the PCG strategy, $g_i^b$ is instantiated as the IoU between the predicted box and the pseudobox, providing an intuitive reflection of the consistency between the pseudobox and the proposal, with values ranging from 0 to 1. In the unsupervised learning component of SMDC-SSOD, the consistency between pseudoboxes and a series of corresponding proposals generated by the student network is used as instance-wise regression loss weights, guiding the student network to pay more attention to high-quality pseudoboxes during training, thereby enabling higher quality boundary box predictions for ships in SAR images.

## IV. Experiment and Results

### A. Dataset

*1) HRSID:* The dataset [43], widely utilized in SAR ship perception tasks [44], [45], [46], [47], comprises 5604 SAR image samples with a total of 16 951 ships, sourced from the European Space Agency's Sentinel-1 satellite and the German TerraSAR-X. The images have an average size of 800 × 800 with spatial resolutions ranging from 0.5 to 3 m and polarization modes including HH, HV, and VV. The dataset encompasses ships in varied environmental conditions, both favorable and adverse sea states, and complex coastal and simple offshore scenes. In our experiments, we randomly allocated 35% of the images to the test set (1961 images) while selecting 1%, 2%, 5%, and 10% of the remaining 65% of images (3643 images) for training, utilizing their labels.

*2) BBox-SSDD:* This dataset was obtained by Zhang et al. [48] on the early SSDD dataset initially established by Li et al. [49], which provides 1160 SAR image samples with an average size of 500 × 500, covering various polarization modes. This dataset contains 2587 ships, exhibiting significant scale variations, from the smallest ship occupying only 20 pixels to the largest ship spanning 55 440 pixels. A total of 35% of the images are designated for the test set, while 65% comprise the training set. During semisupervised training, we randomly obtain 1%, 2%, 5%, and 10% of the images from the training set to utilize their labels during training.

*3) SAR-Ship-Dataset:* The dataset [50] is obtained through the Gaofen-3 satellite and Sentinel-1 satellite and comprises 43 819 images, including multiple sources and modes. Each image has dimensions of 256 × 256, with resolutions ranging from 3 to 25 m. Similar to the partitioning method used for HRSID and BBox-SSDD, we also allocated 65% of the SAR-Ship-Dataset for training, with the remainder designated for testing, where 1%, 2%, 5%, and 10% of the images in the training set were chosen to utilize their labels in the semisupervised learning process.

### B. Implementation Details

The proposed SMDC-SSOD is implemented using PyTorch, with the GPU being the NVIDIA Tesla A100. Both the teacher and student networks are based on Faster R-CNN, utilizing a pretrained ResNet-50 from ImageNet as the backbone network and optimized using stochastic gradient descent with a learning

TABLE I
QUALITATIVE DETECTION RESULTS OF VARIOUS SSOD METHODS ON HRSID

| Percentage | Method | AP | $AP_{50}$ | $AP_{75}$ | $AP_S$ | $AP_M$ | $AP_L$ |
|---|---|---|---|---|---|---|---|
| 1% | Supervised baseline | 38.6 | 62.3 | 42.6 | 41.0 | 34.0 | 0 |
| | Unbiased Teacher [17] | 39.2 | 73.5 | 37.9 | 43.3 | 24.0 | 0 |
| | Soft Teacher [19] | 42.9 | 70.3 | 46.0 | 45.4 | 40.3 | 2.1 |
| | PseCo [20] | 43.5 | 71.7 | 46.1 | 46.0 | 40.0 | 2.6 |
| | ASTOD [18] | 42.2 | 70.1 | 45.4 | 45.7 | 38.2 | 2.3 |
| | MixTeacher [22] | 43.7 | 73.2 | 45.9 | 46.1 | 37.8 | **4.0** |
| | SMDC-SSOD | **44.8** | **74.2** | **50.7** | **46.5** | **43.9** | 2.4 |
| 2% | Supervised baseline | 42.4 | 66.3 | 47.9 | 45.3 | 33.1 | 1.2 |
| | Unbiased Teacher [17] | 44.8 | 70.8 | 50.9 | 47.9 | 35.1 | 2.5 |
| | Soft Teacher [19] | 46.4 | 72.2 | 53.8 | 49.4 | 37.1 | 3.7 |
| | PseCo [20] | 46.9 | 71.8 | 53.9 | 50.3 | **38.3** | 3.1 |
| | ASTOD [18] | 46.5 | 72.7 | 53.3 | 48.0 | 36.2 | 4.2 |
| | MixTeacher [22] | 46.9 | 73.6 | 53.8 | 50.3 | 38.0 | 4.7 |
| | SMDC-SSOD | **48.2** | **76.2** | **54.5** | **51.9** | 37.7 | **5.8** |
| 5% | Supervised baseline | 47.8 | 74.6 | 54.1 | 50.1 | 43.3 | 2.8 |
| | Unbiased Teacher [17] | 48.4 | 74.1 | 55.5 | 50.8 | 44.4 | 5.0 |
| | Soft Teacher [19] | 49.7 | 75.2 | 57.2 | 51.5 | 46.5 | 6.2 |
| | PseCo [20] | 49.8 | 75.8 | 56.4 | 52.2 | 46.4 | 8.2 |
| | ASTOD [18] | 49.2 | 75.9 | 56.5 | 51.3 | 46.0 | 5.5 |
| | MixTeacher [22] | 51.4 | 78.8 | **58.3** | 53.8 | 46.3 | 7.4 |
| | SMDC-SSOD | **52.3** | **80.0** | 58.2 | **54.9** | **48.2** | **8.9** |
| 10% | Supervised baseline | 50.1 | 75.9 | 57.5 | 52.2 | 47.9 | 4.7 |
| | Unbiased Teacher [17] | 50.9 | 76.2 | 58.2 | 53.0 | 48.7 | 6.8 |
| | Soft Teacher [19] | 51.5 | 76.3 | 58.1 | 53.5 | 48.6 | **10.8** |
| | PseCo [20] | 51.7 | 78.7 | 58.3 | 53.3 | 50.2 | 8.0 |
| | ASTOD [18] | 51.2 | 77.0 | 57.8 | 52.9 | 51.0 | 9.4 |
| | MixTeacher [22] | 51.7 | 79.1 | 58.4 | 53.7 | 49.4 | 7.3 |
| | SMDC-SSOD | **53.4** | **80.5** | **60.2** | **55.6** | **51.7** | 9.6 |

The bold values indicate the best performance under their corresponding label annotation percentages.

rate of 2.5e-3, momentum of 0.9, weight decay of 1e-4, and a total of 18k iterations, reducing the learning rate to one-tenth at 11k and 16k iterations. Weak and strong data augmentations in the teacher and student networks are consistent with Soft Teacher. The neighborhood size in the CSFM scheme is set to $5 \times 5$, and the unsupervised learning component only involves the intrapyramid feature cross-scale mixing. The threshold $\Delta S$ for the pseudolabel screening indicator $\varphi$ is set to 0.1. During training, teacher network predictions with confidence scores greater than 0.9 are directly considered pseudolabels, and pseudolabels are selected using the SCCG strategy from teacher network predictions with confidence scores ranging from 0.7 to 0.9.

### C. Evaluation Metrics

This study employs the evaluation metrics from the MS-COCO dataset [51] to validate detection performance, primarily including AP, $AP_{50}$, $AP_{75}$, $AP_S$, $AP_M$, and $AP_L$. Here, AP is the average of multiple average precisions under different IoU thresholds from 0.5 to 0.95 at intervals of 0.05. $AP_{50}$ and $AP_{75}$ represent the average precision at IoU thresholds of 0.5 and 0.75, respectively. Furthermore, $AP_S$, $AP_M$, and $AP_L$ represent the detection performance for ships of varying sizes, where $AP_S$ applies to small ships (area $< 32 \times 32$), $AP_M$ corresponds to medium-sized ships ($32 \times 32 \leq$ area $< 96 \times 96$), and $AP_L$ focuses on larger targets (area $\geq 96 \times 96$).

### D. Comparative Experimental Results

In this section, we present the detection performance of SMDC-SSOD on the HRSID, BBox-SSDD, and SAR-Ship-Dataset with labeled data proportions of 1%, 2%, 5%, and 10%, comparing it against supervised baseline and various state-of-the-art SSOD methods such as Unbiased Teacher [17], Soft Teacher [19], PseCo [20], ASTOD [18] and MixTeacher [22] to comprehensively validate the effectiveness of SMDC-SSOD in detecting ships in SAR images. The network architecture of the supervised baseline is the standard Faster R-CNN [2], comprising ResNet, FPN, RPN, classification head, and regression head, and it adopts the same data augmentation methods as those in the student network within SMDC-SSOD.

*1) Experimental Results on HRSID:* Table I illustrates the detection performance of SMDC-SSOD and its comparative methods on HRSID with labeled data proportions of 1%, 2%, 5%, and 10%. Because the supervised baseline merely utilizes labeled data for training without effectively utilizing unlabeled data, its overall detection performance is weaker than SSOD

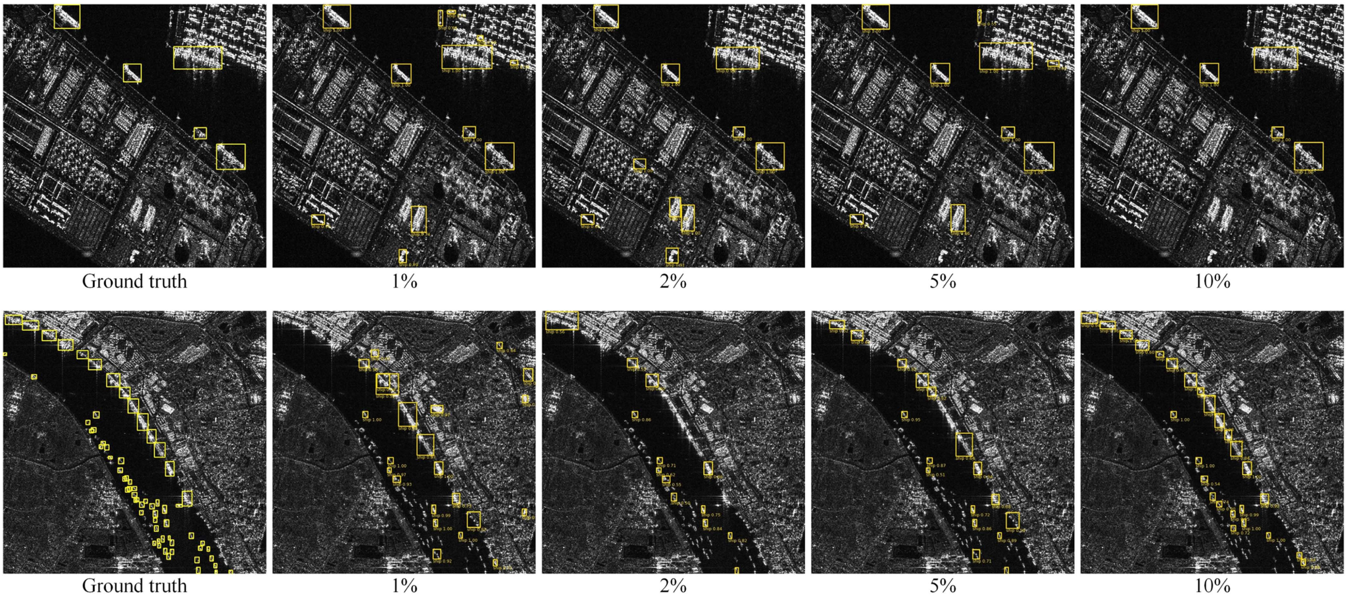


Fig. 5. Qualitative results of SMDC-SSOD on HRSID at various labeled data proportions.

methods. At a labeled data proportion of 1%, SMDC-SSOD achieves a commendable performance of 44.8 AP, surpassing the multistage trained SSOD method Unbiased Teacher and ASTOD by 5.6 AP and 2.6 AP, respectively. It also outperforms various end-to-end trained SSOD methods. Therefore, the proposed SMDC-SSOD can effectively perceive ships in SAR images. Furthermore, as the labeled data proportion gradually increases, the detection performance of all methods also improves. At a labeled data proportion of 5%, SMDC-SSOD achieves an AP of 52.3, not only surpassing the six comparative methods at the same labeled data proportion but also outperforming the six comparative methods at a labeled data proportion of 10%, further demonstrating the outstanding performance of the proposed method in ship detection in SAR images. Overall, our SMDC-SSOD consistently outperforms other semisupervised detection methods at labeled data proportions of 1%, 2%, 5%, and 10%, demonstrating its strong ship perception capabilities under conditions of limited labeled data.

In addition to quantitative detection results, we also visualize the SAR ship detection results on HRSID. Fig. 5 reflects the detection performance of SMDC-SSOD at different labeled data proportions. It can be observed that at a labeled data proportion of 1%, some background in the images is falsely recognized as ships, resulting in numerous false alarms (false positives). In addition, SMDC-SSOD exhibits a significant number of missed detections (false negatives) for small-scale ships in the lower portion of Fig. 5. As the labeled data proportion gradually increases, the occurrences of false alarms and missed detections also improve. Under a labeled data proportion of 10%, SMDC-SSOD shows no false alarms in the upper portion of Fig. 5 and a noticeable improvement in missed detections for the smaller ships in the lower portion of Fig. 5. Fig. 6 illustrates the qualitative experimental results of SMDC-SSOD compared to semisupervised detection methods at a labeled data proportion of 10%. Due to space constraints, we have only visualized the quantitative results of the supervised baseline and the comparative methods with good quantitative performance, PseCo and MixTeacher, to compare them with our proposed SMDC-SSOD qualitatively. While the supervised baseline, PseCo, and MixTeacher can detect most ships, they exhibit a certain degree of false alarms and missed detections, particularly evident in the more complex scenes in the first three columns. In comparison, although our SMDC-SSOD also shows a few false alarms and missed detections, its detection results better align with the ground truth in the first row, indicating that the proposed SMDC-SSOD demonstrates stronger ship perception capabilities in SAR images.

2) *Experimental Results on BBox-SSDD:* Table II reflects the detection performance of various methods on BBox-SSDD at labeled data proportions of 1%, 2%, 5%, and 10%. Because the supervised baseline only utilizes labeled data for training without effectively using unlabeled data, when the labeled data is at 1%, the supervised baseline achieves an AP of only 28.8, significantly lower than the AP of SMDC-SSOD. Moreover, the AP values of semisupervised methods such as Unbiased Teacher and MixTeacher are also lower than that of SMDC-SSOD, indicating the superior SAR ship detection capability of the proposed SMDC-SSOD compared to the comparative methods. Similar to the HRSID, the detection performance of all methods on BBox-SSDD also improves as the proportion of labeled data increases. When the proportion of labeled annotations reaches 10%, SMDC-SSOD achieves an AP of 61.6, which is 6.2 higher than the supervised baseline. In addition, it surpasses the semisupervised methods such as Unbiased Teacher, Soft Teacher, PseCo, ASTOD, and MixTeacher by 5.1, 3.3, 2.1, 3.4, and 1.0, respectively, further demonstrating the outstanding performance

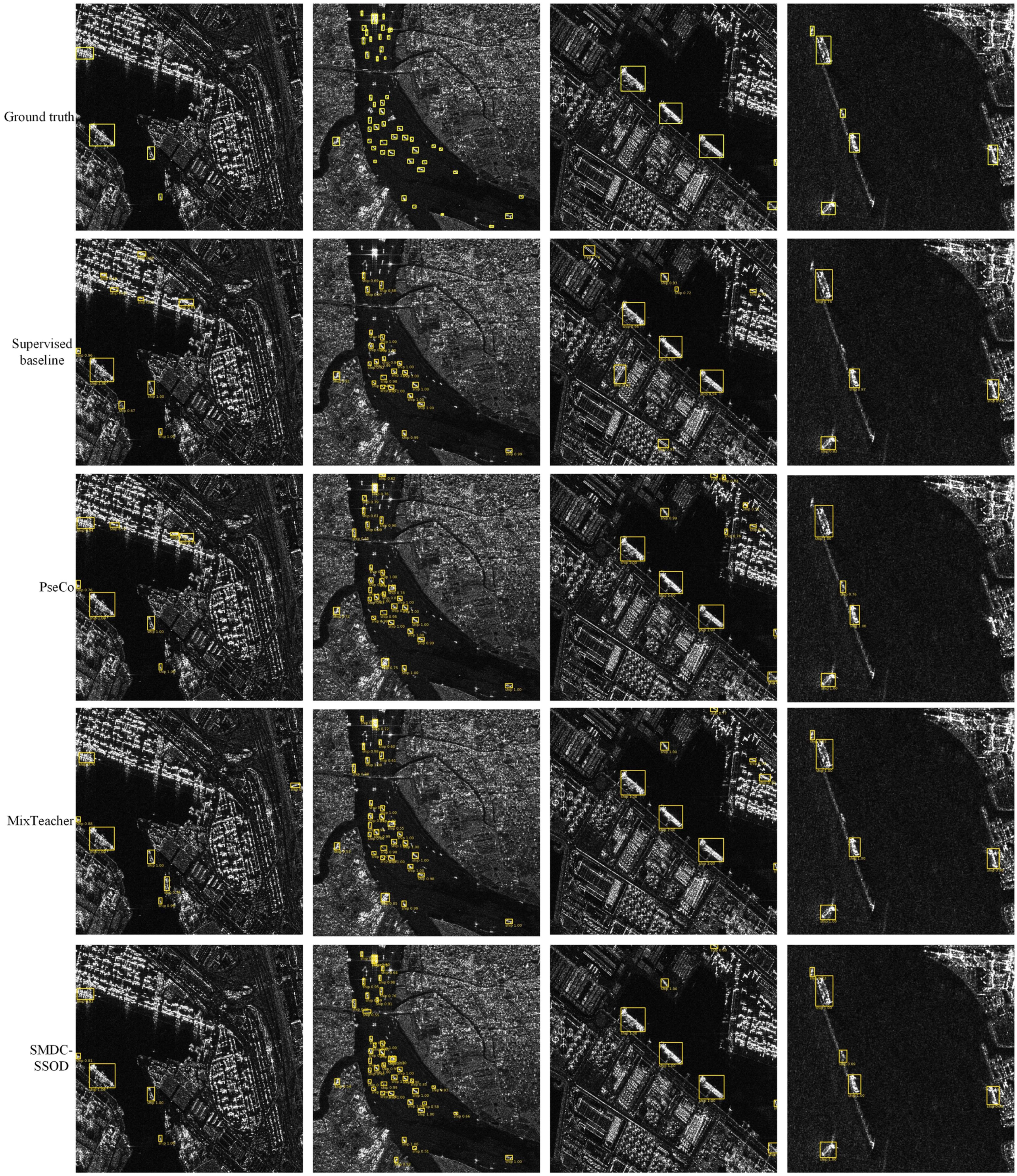


Fig. 6. Qualitative results comparing various SSOD methods on HRSID.

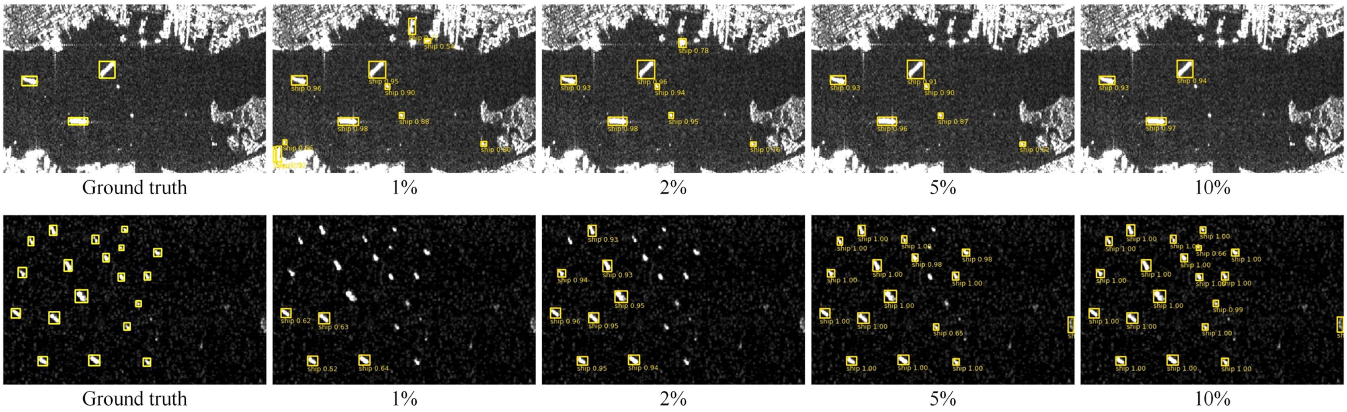
of the proposed SMDC-SSOD in SAR ship object detection. Furthermore, the fluctuation range of APL is relatively large due to the relatively small scale of BBox-SSDD and the low proportion of large ships (approximately 2%). The proposed SMDC-SSOD exhibits better segmentation performance at various labeled data proportions than the comparative methods due to its cross-scale feature fusion scheme and two consistency-guided strategies.

Fig. 7 also demonstrates the detection performance of SMDC-SSOD at various labeled data proportions. At a labeled data proportion of 1%, there are numerous false alarms in the upper portion, incorrectly identifying objects in the background as ships, while the lower portion exhibits a significant number of missed detections. As the labeled data increases, the detection performance of SMDC-SSOD gradually improves, with the

TABLE II
QUALITATIVE DETECTION RESULTS OF VARIOUS SSOD METHODS ON BBOX-SSDD

| Percentage | Method | AP | $AP_{50}$ | $AP_{75}$ | $AP_S$ | $AP_M$ | $AP_L$ |
|---|---|---|---|---|---|---|---|
| 1% | Supervised baseline | 28.8 | 57.4 | 22.4 | 38.3 | 8.2 | 0.9 |
| | Unbiased Teacher [17] | 30.0 | 59.1 | 28.1 | 40.0 | 10.1 | 1.5 |
| | Soft Teacher [19] | 30.9 | 58.5 | 29.4 | 39.7 | **13.6** | 9.3 |
| | PseCo [20] | 31.7 | 58.4 | 29.9 | 40.7 | 12.4 | 4.7 |
| | ASTOD [18] | 30.3 | 58.5 | 28.6 | 39.9 | 12.0 | 7.4 |
| | MixTeacher [22] | 32.0 | 59.1 | 31.4 | 41.3 | 11.9 | **12.8** |
| | SMDC-SSOD | **33.7** | **60.9** | **33.3** | **43.5** | 12.1 | 8.4 |
| 2% | Supervised baseline | 43.6 | 71.8 | 49.2 | 50.0 | 18.3 | 13.0 |
| | Unbiased Teacher [17] | 45.4 | 78.6 | 51.7 | 50.4 | 32.8 | 15.3 |
| | Soft Teacher [19] | 46.5 | 77.5 | 52.5 | 50.9 | 34.0 | 22.7 |
| | PseCo [20] | 48.3 | 80.4 | 54.6 | 53.7 | 35.3 | 17.7 |
| | ASTOD [18] | 47.0 | 78.2 | 52.6 | 51.5 | 33.7 | 19.4 |
| | MixTeacher [22] | 49.6 | **82.9** | 55.5 | 54.8 | 36.9 | 9.4 |
| | SMDC-SSOD | **50.8** | 82.4 | **57.1** | **55.0** | **39.7** | **25.1** |
| 5% | Supervised baseline | 49.7 | 83.0 | 54.1 | 54.2 | 39.4 | 2.3 |
| | Unbiased Teacher [17] | 51.5 | 84.3 | 58.7 | 55.6 | 49.3 | 15.2 |
| | Soft Teacher [19] | 53.6 | 84.1 | 63.4 | 55.0 | 52.0 | 26.8 |
| | PseCo [20] | 55.2 | 85.4 | 65.7 | **59.7** | 53.4 | 20.2 |
| | ASTOD [18] | 52.8 | 83.9 | 61.5 | 55.4 | 52.3 | 22.6 |
| | MixTeacher [22] | 54.9 | 85.6 | 66.3 | 56.5 | 52.8 | 18.7 |
| | SMDC-SSOD | **56.5** | **87.3** | **68.1** | 59.5 | **54.3** | **30.2** |
| 10% | Supervised baseline | 55.4 | 87.6 | 64.7 | 58.8 | 47.9 | 20.3 |
| | Unbiased Teacher [17] | 56.5 | 88.0 | 68.1 | 60.6 | 48.4 | 8.7 |
| | Soft Teacher [19] | 58.3 | 89.7 | 70.4 | 59.2 | 56.4 | **32.4** |
| | PseCo [20] | 59.5 | 90.8 | 70.7 | 62.1 | 54.2 | 18.5 |
| | ASTOD [18] | 58.2 | 89.2 | 69.7 | 61.4 | 52.6 | 26.4 |
| | MixTeacher [22] | 60.6 | 92.5 | 72.0 | 63.2 | 54.1 | 27.3 |
| | SMDC-SSOD | **61.6** | **93.6** | **73.1** | **64.0** | **58.1** | 31.3 |

The bold values indicate the best performance under their corresponding label annotation percentages.


Ground truth 1% 2% 5% 10%

Ground truth 1% 2% 5% 10%

Fig. 7. Qualitative results of SMDC-SSOD on BBox-SSDD at various labeled data proportions.

occurrences of false alarms and missed detections gradually ameliorating. Fig. 8 reports the qualitative experimental results of semisupervised methods in BBox-SSDD. In the scenes depicted in the first column, the supervised baseline fails to detect ships and exhibits false alarms. Pseco also shows occurrences of false alarms, while MixTeacher, although devoid of false alarms, does not effectively align its detected bounding boxes with the ground truth. In comparison, our proposed SMDC-SSOD exhibits no false alarm occurrences and better aligns with the ground truth in the first row. The supervised baseline shows false alarms and missed detections in the second column, and PseCo and MixTeacher show two false alarms. In contrast, our SMDC-SSOD shows better performance with no missed detections and only one false alarm. In the third column, where numerous

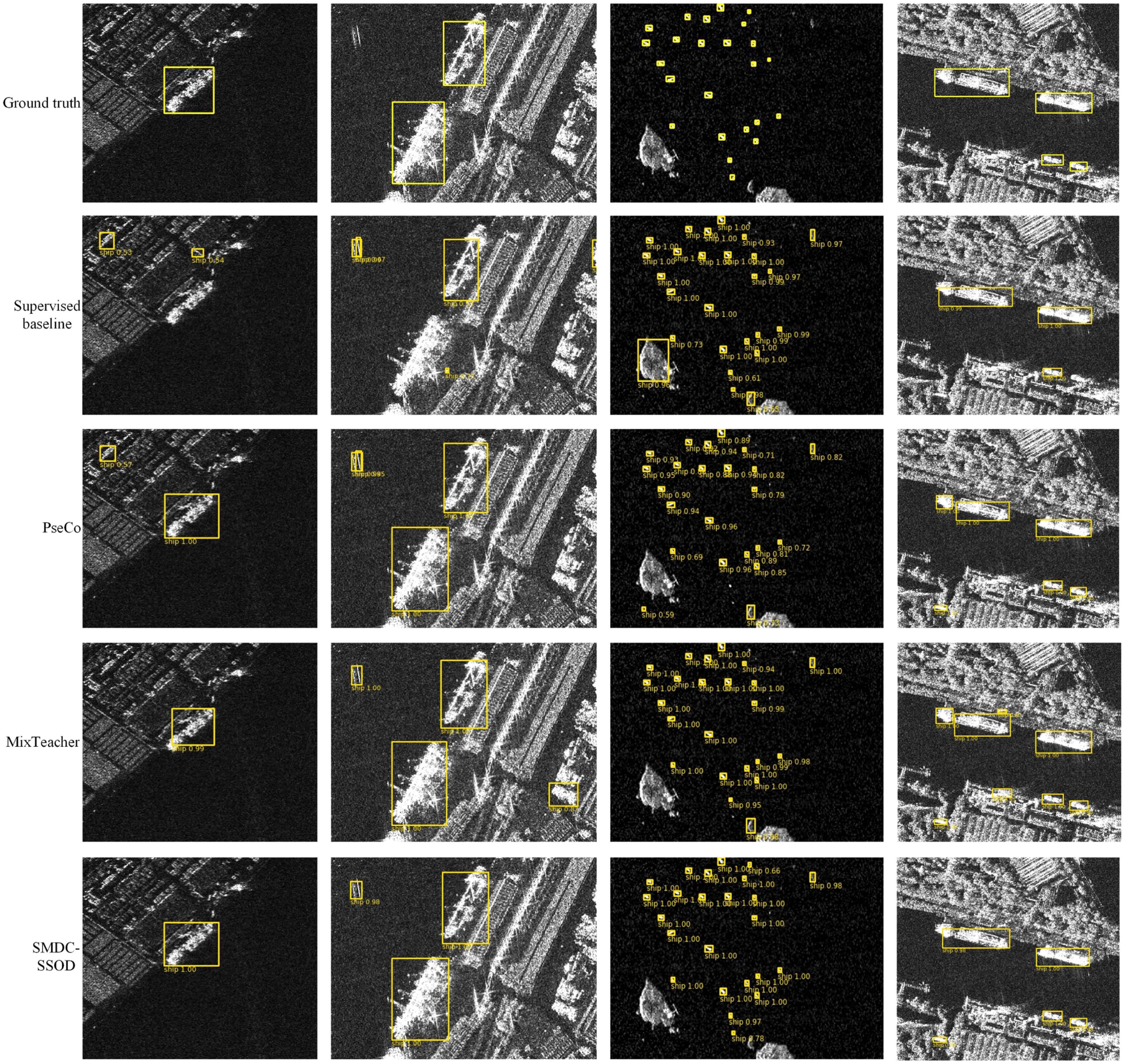


Fig. 8. Qualitative results comparing various SSOD methods on BBox-SSDD.

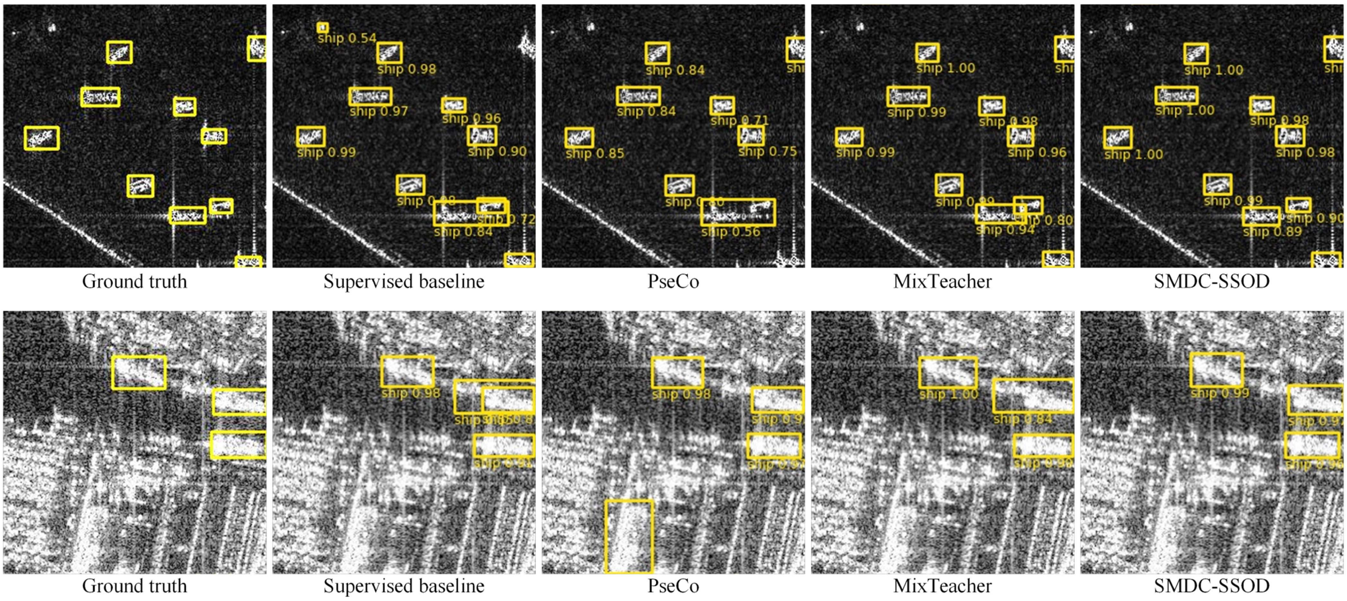
small targets are present, all comparative methods exhibit a certain degree of missed detections. Our SMDC-SSOD can detect most ships except for extremely tiny targets. In the fourth column, the supervised baseline displays missed detections, and PseCo and MixTeacher experience significant false alarms. Despite misidentifying one object in the background as a ship, our proposed SMDC-SSOD outperforms the comparative methods. In summary, our SMDC-SSOD demonstrates a better perception of ships for images in BBox-SSDD.

*3) Experimental Results on SAR-Ship-Dataset:* Table III reports the quantitative detection results on the SAR-Ship-Dataset, showcasing only the AP values to avoid excessive length. It can be observed that as the proportion of annotated data increases, the overall detection performance gradually improves. Due to the inability of the supervised baseline to effectively utilize unlabeled data, its detection performance is weaker than that of all SSOD methods with the same annotation ratio. At a labeled data proportion of 1%, our proposed SMDC-SSOD not only surpasses the supervised baseline by 4.0 in terms of AP but also outperforms the best-performing SSOD method, MixTeacher, by 0.8, highlighting the positive role of CSFM and dual-consistency guidance. When the labeled data accounts for 2%, 5%, and 10%, our SMDC-SSOD also demonstrates satisfactory overall performance, with AP values surpassing the supervised baseline by 3.7, 2.8, and 3.5, respectively, and outperforming other SSOD methods at the same annotation ratios.

TABLE III
QUALITATIVE DETECTION RESULTS OF VARIOUS SSOD METHODS ON SAR-SHIP-DATASET

| Method | 1% | 2% | 5% | 10% |
|---|---|---|---|---|
| Supervised baseline | 40.5 | 43.7 | 45.8 | 47.3 |
| Unbiased Teacher [17] | 42.3 | 44.2 | 46.5 | 47.6 |
| Soft Teacher [19] | 43.4 | 44.0 | 46.7 | 48.5 |
| PseCo [20] | 43.8 | 45.6 | 47.1 | 49.0 |
| ASTOD [18] | 42.9 | 44.9 | 46.9 | 48.2 |
| MixTeacher [22] | 43.7 | 46.5 | 47.4 | 49.4 |
| SMDC-SSOD | **44.5** | **47.4** | **48.6** | **50.8** |

The bold values indicate the best performance.



Ground truth Supervised baseline PseCo MixTeacher SMDC-SSOD

Ground truth Supervised baseline PseCo MixTeacher SMDC-SSOD

Fig. 9. Qualitative results comparing various SSOD methods on SAR-ship-dataset.

Thus, our SMDC-SSOD showcases excellent ship perception capabilities on the SAR-Ship-Dataset.

We also present the qualitative experimental results on the SAR-Ship-Dataset in Fig. 9. In the first row, the supervised baseline exhibits both missed detections and false alarms, and PseCo shows instances of missed detections. Although MixTeacher does not display missed detections or false alarms, its detected bounding boxes for some ships deviate significantly from the ground truth. Therefore, the detection of these three comparative methods is unsatisfactory. In contrast, our SMDC-SSOD avoids missed detections and false alarms and demonstrates a higher degree of alignment between the detected bounding boxes and the ground truth. In the second row with intense speckle noise, the number of ships detected by the supervised baseline and PseCo does not align with the ground truth. The alignment of MixTeacher's detected bounding boxes with the ground truth is lower compared to our proposed SMDC-SSOD. Through the analysis above, it is evident that our proposed SMDC-SSOD exhibits superior detection performance.

*4) Model Complexity Comparison:* Table IV reports the floating point operations per second (FLOPs) of the proposed SMDC-SSOD and comparative methods to reflect the complexity of each model. Since Unbiased Teacher, Soft Teacher, PseCo, and MixTeacher are all based on the design of Faster R-CNN, their overall structures are similar, thus resulting in similar FLOPs (minor discrepancies might arise from specific implementation differences). Furthermore, due to the slight adjustments to the model structure in the intrapyramid feature cross-scale mixing part of the CSFM scheme within SMDC-SSOD, its FLOPs have a slightly higher value compared to Unbiased Teacher, Soft Teacher, PseCo, and MixTeacher, but this increase is minimal. Combined with the qualitative detection

TABLE IV
MODEL COMPLEXITY COMPARISON

| Method | FLOPs/G |
|---|---|
| Unbiased Teacher [17] | 204.1 |
| Soft Teacher [19] | 202.3 |
| PseCo [20] | 203.2 |
| MixTeacher [22] | 202.3 |
| SMDC-SSOD | 205.3 |

TABLE V
ABLATION EXPERIMENTS OF CORE COMPONENTS ON HISID

| Method | CSFM | SVCG | PCG | AP | $AP_{50}$ | $AP_{75}$ | $AP_S$ | $AP_M$ | $AP_L$ | FPS |
|---|---|---|---|---|---|---|---|---|---|---|
| Baseline | | | | 51.0 | 76.6 | 57.8 | 52.9 | 48.4 | 8.2 | 26.2 |
| Components | √ | | | 52.3 | 78.9 | 58.9 | 54.4 | 50.1 | 9.5 | 25.5 |
| | | | √ | 51.7 | 77.2 | 58.3 | 52.7 | 49.5 | 7.3 | 26.2 |
| | √ | | √ | 52.8 | 79.6 | 59.8 | 55.0 | 50.8 | 8.8 | 25.5 |
| | √ | √ | | 53.0 | 79.8 | 59.6 | 54.9 | 50.9 | 10.0 | 25.4 |
| SMDC-SSOD | √ | √ | √ | 53.4 | 80.5 | 60.2 | 55.6 | 51.7 | 9.6 | 25.4 |

TABLE VI
ABLATION EXPERIMENTS OF CORE COMPONENTS ON BBOX-SSDD

| Method | CSFM | SCCG | PCG | AP | $AP_{50}$ | $AP_{75}$ | $AP_S$ | $AP_M$ | $AP_L$ | FPS |
|---|---|---|---|---|---|---|---|---|---|---|
| Baseline | | | | 57.7 | 89.0 | 69.3 | 58.9 | 54.1 | 17.0 | 32.4 |
| Components | √ | | | 59.9 | 91.6 | 71.2 | 61.8 | 56.4 | 26.7 | 31.5 |
| | | | √ | 58.6 | 88.9 | 70.7 | 59.8 | 54.4 | 12.4 | 32.3 |
| | √ | | √ | 60.5 | 92.3 | 72.4 | 62.9 | 53.8 | 20.5 | 31.5 |
| | √ | √ | | 60.8 | 93.1 | 72.2 | 62.7 | 57.5 | 15.6 | 31.5 |
| SMDC-SSOD | √ | √ | √ | 61.6 | 93.6 | 73.1 | 64.0 | 58.1 | 31.3 | 31.4 |

results, it is evident that our SMDC-SSOD achieves high-quality detection performance without significantly increasing complexity.

### E. Ablation Experiments

*1) Analysis of Core Components in SMDC-SSOD:* In order to validate the contribution of the CSFM scheme, SCCG strategy, and PCG strategy in the proposed SMDC-SSOD for SAR ship detection, detailed ablation experiments were conducted, as given in Tables V and VI. In the experiment, SMDC-SSOD without the CSFM scheme, SCCG strategy, and PCG strategy were considered the baseline, and all experiments in this section were performed with a labeled data proportion of 10%. It can be observed that with the introduction of the CSFM scheme, SMDC-SSOD achieved a certain improvement in AP values on HRSID and SSDD, indicating that cross-scale feature fusion plays a positive role in enhancing the perception capability of ships in SAR images. Benefiting from the SCCG strategy, designed to filter pseudolabels based on confidence score variations at different scales, and the PCG strategy, which utilizes the consistency of proposals generated by the student network to reflect the quality of pseudolabels further, the detection performance of SMDC-SSOD on HRSID and SSDD was enhanced upon the introduction of the SCCG and PCG strategies. Overall, the detection performance of SMDC-SSOD, incorporating the CSFM scheme, SCCG strategy, and PCG strategy, reached an optimal level, further demonstrating the positive significance of the three designed components for SAR ship detection.

TABLE VII
ABLATION EXPERIMENTS OF CSFM IN THE CSFM SCHEME

| Dataset | Operation | AP | $AP_{50}$ | $AP_{75}$ | $AP_S$ | $AP_M$ | $AP_L$ |
|---|---|---|---|---|---|---|---|
| HRSID | Baseline | 51.0 | 76.6 | 57.8 | 52.9 | 48.4 | 8.2 |
| | + Inter | 51.9 | 78.0 | 58.6 | 53.9 | 49.3 | 7.9 |
| | + Intra | 51.6 | 77.1 | 58.5 | 52.8 | 48.8 | 8.7 |
| | + All | 52.3 | 78.9 | 58.9 | 54.4 | 50.1 | 9.5 |
| BBox-SSDD | Baseline | 57.7 | 89.0 | 69.3 | 58.9 | 54.1 | 17.0 |
| | + Inter | 59.0 | 90.5 | 70.4 | 60.6 | 55.1 | 16.2 |
| | + Intra | 58.5 | 90.4 | 70.3 | 60.3 | 55.0 | 29.1 |
| | + All | 59.9 | 91.6 | 71.2 | 61.8 | 56.4 | 26.7 |

Note: "Inter" and "Intra" denote inter-pyramid and intra-pyramid feature cross-scale mixings, respectively. "All" signifies the presence of both types of feature cross-scale mixings, thus constituting a comprehensive CSFM scheme.

In addition to the detection performance metrics, we reported the frames per second (FPS) in Tables V and VI to comprehensively reflect the model's performance. Regarding inference speed, introducing the CSFM scheme led to a slight decrease in FPS for SMDC-SSOD. As the SCCG and PCG strategies are only used during the training process, their introduction did not significantly alter the inference speed of SMDC-SSOD, with only minor fluctuations in FPS. Overall, our CSFM scheme, SCCG strategy, and PCG strategy had minimal impact on the model's inference speed.

*2) Analysis of Feature Cross-Scale Mixing in the CSFM Scheme:* The CSFM scheme can achieve CSFM from both inter-pyramid and intrapyramid aspects. Further ablation experiments were conducted on these two components of CSFM to verify their impact on ship detection performance. The experimental results are presented in Table VII, with the baseline identical to that in Tables V and VI. On HRSID, the introduction of CSFM between and within pyramids resulted in a certain improvement in AP for SMDC-SSOD, indicating their ability to facilitate more accurate ship detection in SAR images. In BBox-SSDD, with the introduction of interpyramid feature cross-scale mixing, SMDC-SSOD's AP increased by 1.3. When intrapyramid feature cross-scale mixing was added, SMDC-SSOD's AP improved by 0.8 compared to the baseline. With both introduced, SMDC-SSOD's AP increased by 2.2 compared to the baseline.

TABLE VIII
ABLATION EXPERIMENTS OF THE NEIGHBORHOOD SIZE IN THE CSFM SCHEME

| Dataset | Size | AP | $AP_{50}$ | $AP_{75}$ | $AP_S$ | $AP_M$ | $AP_L$ |
|---|---|---|---|---|---|---|---|
| HRSID | 3×3 | 52.0 | 78.5 | 58.7 | 54.0 | 49.6 | 8.9 |
| | 5×5 | 52.3 | 78.9 | 58.9 | 54.4 | 50.1 | 9.5 |
| | 7×7 | 52.2 | 78.7 | 59.1 | 54.3 | 49.9 | 9.8 |
| BBox-SSDD | 3×3 | 59.4 | 91.0 | 70.8 | 61.2 | 55.7 | 19.3 |
| | 5×5 | 59.9 | 91.6 | 71.2 | 61.8 | 56.4 | 26.7 |
| | 7×7 | 60.1 | 91.4 | 71.2 | 61.5 | 56.1 | 30.3 |

TABLE IX
ABLATION EXPERIMENTS OF THE THRESHOLD FOR PSEUDOLABEL FILTERING METRICS IN THE SCCG STRATEGY

| Dataset | None | $\varphi$ =0.05 | $\varphi$ =0.1 | $\varphi$ =0.15 | $\varphi$ =0.2 |
|---|---|---|---|---|---|
| HRSID | 52.3 | 52.7 | 53.0 | 52.5 | 52.4 |
| BBox-SSDD | 59.9 | 60.4 | 60.8 | 60.1 | 59.9 |

Note: "None" indicates the absence of the SCCG strategy.

Therefore, both interpyramid and intrapyramid CSFMs have a positive impact on SAR ship detection.

*3) Analysis of Neighborhood Size in the CSFM Scheme:* Since the neighborhood size in the kernel prediction of the CSFM scheme may influence the detection performance of the model, we conducted ablation experiments, and the results are documented in Table VIII. It can be observed that when the neighborhood size in the CSFM scheme is set to 3 × 3, the AP values on HRSID and BBox-SSDD are 52.0 and 59.4, respectively, which are lower compared to the detection performance when the neighborhood size is set to 5 × 5 and 7 × 7. When the neighborhood size is 5 × 5 and 7 × 7, the proposed method achieves an AP of 52.3 and 52.2 on HRSID, respectively, with a minimal difference. For BBox-SSDD, the detection performance is also quite similar when the neighborhood size is 5 × 5 and 7 × 7. Considering that a larger neighborhood size would increase computational complexity, we choose 5 × 5 as the default setting.

*4) Threshold Analysis of Pseudolabel Filtering Metrics in the SCCG Strategy:* Considering that the threshold of pseudolabel filtering metrics in the SCCG strategy may affect pseudolabel filtering, we conducted ablation experiments for it, as given in Table IX. It is evident that without the SCCG strategy, SMDC-SSOD exhibited the poorest detection performance. When the threshold value was set at 0.1, SMDC-SSOD achieved the best detection performance on HRSID and BBox-SSDD, with AP values of 53.0 and 60.8, respectively, indicating that this threshold effectively distinguished valuable predictions for the teacher network, thereby guiding the student network to obtain better predictions. Setting the threshold at 0.15 and 0.2 led to overly stringent filtering by the teacher network, suppressing valuable predictions, thus yielding less noticeable improvements in SMDC-SSOD's detection performance. Overall, a threshold of 0.1 for the pseudolabel filtering metric in the SCCG strategy yielded the best perception of ships in SAR images by SMDC-SSOD, leading us to set this threshold at 0.1 in our study.

## V. DISCUSSION

In this study, we focused on the strong dependence on data labels, the significant impact of pseudolabels, and inherent challenges, such as multiscale ship feature disparities and indistinct small-sized ships in SAR ship detection tasks, proposing the SAR ship SSOD method, SMDC-SSOD. This method is primarily based on TSF, encompassing three core components: the CSFM scheme, SCCG strategy, and PCG strategy, enabling end-to-end semisupervised learning from limited annotated data and achieving better SAR ship detection performance compared to state-of-the-art SSOD methods. This work provides a low-cost and high-performance solution for ship detection tasks in SAR images.

While our method has demonstrated favorable results, it also possesses certain limitations. First, although our method significantly reduces dependence on data labeling, its detection performance is not ideal when faced with minimal annotated data (e.g., 1% labeled data). Second, our method can only achieve bounding box-level ship detection and cannot achieve more fine-grained pixel-level perception. Finally, this study primarily focuses on addressing the dependence on data labels. However, it does not emphasize model complexity and inference speed, so the proposed method does not present significant advantages in complexity and inference speed, potentially limiting its application in scenarios requiring real-time SAR ship detection.

In future research, we aim to further explore SAR ship detection methods in scenarios with minimal annotated data, considering the introduction of self-supervised learning or transfer learning to enhance detection performance under extremely limited annotated data conditions. We will also focus on semisupervised segmentation methods for SAR detection to achieve low-cost and fine-grained perception of SAR ship targets. Furthermore, we aim to explore SSOD methods for SAR ship detection with lower complexity and faster detection speeds to better adapt to real-time requirements and enhance its practical value. In addition, we are considering researching noncomputer vision SAR ship detection methods to provide further support for practical applications in real-world scenarios.

## VI. CONCLUSION

In this article, the SMDC-SSOD method is proposed. This method primarily encompasses three core components: the CSFM scheme, the SCCG strategy, and the PCG strategy. The CSFM scheme enhances the network's adaptability to scale variations and indistinct small-sized ships by leveraging interpyramid and intrapyramid feature cross-scale mixings. The SCCG strategy utilizes variations in confidence scores at different scales to design pseudolabel filtering metrics, aiding the teacher network in filtering more valuable pseudolabels in predictions and providing more precise guidance to the student network. The PCG strategy reflects the localization quality of pseudolabels based on the consistency of proposals generated by the student network, guiding the student network to make high-quality boundary predictions. Experimental results on HRSID, BBox-SSDD, and SAR-Ship-Dataset demonstrate that SMDC-SSOD significantly reduces the dependence on data labels for SAR ship detection. It achieves optimal SAR ship detection performance across different labeled data proportions (i.e., 1%, 2%, 5%, and 10%) compared to state-of-the-art SSOD methods. Ablation studies further validate the effectiveness of each component in the proposed method.

## References

[1] R. M. Asiyabi, A. Ghorbanian, S. N. Tameh, M. Amani, S. Jin, and A. Mohammadzadeh, "Synthetic aperture radar (SAR) for ocean: A review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 9106–9138, 2023, doi: 10.1109/JSTARS.2023.3310363.

[2] 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.

[3] 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.

[4] 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.

[5] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 213–229.

[6] L. Zhang, Y. Liu, W. Zhao, X. Wang, G. Li, and Y. He, "Frequency-adaptive learning for SAR ship detection in clutter scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5215514, doi: 10.1109/TGRS.2023.3249349.

[7] M. Sun, Y. Li, X. Chen, Y. Zhou, J. Niu, and J. Zhu, "A fast and accurate small target detection algorithm based on feature fusion and cross-layer connection network for the SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8969–8981, 2023, doi: 10.1109/JSTARS.2023.3316309.

[8] Z. Huang, Y. Liu, X. Yao, J. Ren, and J. Han, "Uncertainty exploration: Toward explainable SAR target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4202314, doi: 10.1109/TGRS.2023.3247898.

[9] H. Wan et al., "Orientation detector for ship targets in SAR images based on semantic flow feature alignment and Gaussian label matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5218616, doi: 10.1109/TGRS.2023.3323143.

[10] L. Zhou, Z. Wan, S. Zhao, H. Han, and Y. Liu, "BFEA: A SAR ship detection model based on attention mechanism and multiscale feature fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 11163–11177, 2024, doi: 10.1109/JSTARS.2024.3408339.

[11] Y. Li, W. Liu, and R. Qi, "Multilevel pyramid feature extraction and task decoupling network for SAR ship detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 3560–3570, 2024, doi: 10.1109/JSTARS.2023.3347454.

[12] L. Ying, D. Miao, and Z. Zhang, "A robust one-stage detector for SAR ship detection with sequential three-way decisions and multi-granularity," *Inf. Sci.*, vol. 667, 2024, Art. no. 120436, doi: 10.1016/j.ins.2024.120436.

[13] R. Yasarla, V. A. Sindagi, and V. M. Patel, "Semi-supervised image deraining using Gaussian processes," *IEEE Trans. Image Process.*, vol. 30, pp. 6570–6582, 2021, doi: 10.1109/TIP.2021.3096323.

[14] C. Qin, L. Wang, Q. Ma, Y. Yin, H. Wang, and Y. Fu, "Semi-supervised domain adaptive structure learning," *IEEE Trans. Image Process.*, vol. 31, pp. 7179–7190, 2022, doi: 10.1109/TIP.2022.3215889.

[15] L. Wang and K.-J. Yoon, "Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 3048–3068, Jun. 2022, doi: 10.1109/TPAMI.2021.3055564.

[16] K. Sohn, Z. Zhang, C.-L. Li, H. Zhang, C.-Y. Lee, and T. Pfister, "A simple semi-supervised learning framework for object detection," 2020, *arXiv: 2005.04757*.

[17] Y.-C. Liu et al., "Unbiased teacher for semi-supervised object detection," in *Proc. Int. Conf. Learn. Representations*, 2021.

[18] R. Vandeghen, G. Louppe, and M. Van Droogenbroeck, "Adaptive self-training for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2023, pp. 914–923, doi: 10.1109/ICCVW60793.2023.00098.

[19] M. Xu et al., "End-to-end semi-supervised object detection with soft teacher," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 3040–3049, doi: 10.1109/ICCV48922.2021.00305.

[20] G. Li, X. Li, Y. Wang, Y. Wu, D. Liang, and S. Zhang, "PseCo: Pseudo labeling and consistency training for semi-supervised object detection," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 457–472, doi: 10.1007/978-3-031-20077-9_27.

[21] B. Chen et al., "Label matching semi-supervised object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 14361–14370, doi: 10.1109/CVPR52688.2022.01398.

[22] L. Liu et al., "MixTeacher: Mining promising labels with mixed scale teacher for semi-supervised object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 7370–7379, doi: 10.1109/CVPR52729.2023.00712.

[23] W. Kong, S. Liu, M. Xu, M. Yasir, D. Wang, and W. Liu, "Lightweight algorithm for multi-scale ship detection based on high-resolution SAR images," *Int. J. Remote Sens.*, vol. 44, no. 4, pp. 1390–1415, Feb. 2023, doi: 10.1080/01431161.2023.2182652.

[24] S. Liu et al., "A mixed-scale self-distillation network for accurate ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 9843–9857, 2023, doi: 10.1109/JSTARS.2023.3324496.

[25] M. Chen et al., "Semantic attention and structured model for weakly supervised instance segmentation in optical and SAR remote sensing imagery," *Remote Sens.*, vol. 15, no. 21, Nov. 2023, Art. no. 5201, doi: 10.3390/rs15215201.

[26] 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, 2023, doi: 10.1109/JSTARS.2022.3230859.

[27] C. Chen, Y. Zhang, R. Hu, and Y. Yu, "A lightweight SAR ship detector using end-to-end image preprocessing network and channel feature guided spatial pyramid pooling," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024, Art. no. 4003605, doi: 10.1109/LGRS.2024.3358957.

[28] H. Guo and D. Gu, "Closely arranged inshore ship detection using a bi-directional attention feature pyramid network," *Int. J. Remote Sens.*, vol. 44, no. 22, pp. 7106–7125, Nov. 2023, doi: 10.1080/01431161.2023.2277166.

[29] H. Wang, T. Jia, B. Ma, Q. Wang, and W. Zuo, "Fully cascade consistency learning for one-stage object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 10, pp. 5986–5998, Oct. 2023, doi: 10.1109/TCSVT.2023.3263557.

[30] Z. Wang, W. Zhu, W. Zhao, and L. Xu, "Balanced one-stage object detection by enhancing the effect of positive samples," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 8, pp. 4011–4026, Aug. 2023, doi: 10.1109/TCSVT.2023.3237826.

[31] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.

[32] D. Jia et al., "DETRs with hybrid matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 19702–19712, doi: 10.1109/CVPR52729.2023.01887.

[33] L. M. Novak, M. C. Burl, and W. W. Irving, "Optimal polarimetric processing for enhanced target detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 234–244, Jan. 1993, doi: 10.1109/7.249129.

[34] A. Renga, M. D. Graziano, and A. Moccia, "Segmentation of marine SAR images by sublook analysis and application to sea traffic monitoring," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1463–1477, Mar. 2019, doi: 10.1109/TGRS.2018.2866934.

[35] H. Yang, Z. Cao, Z. Cui, and Y. Pi, "Saliency detection of targets in polarimetric SAR images based on globally weighted perturbation filters," *ISPRS J. Photogrammetry Remote Sens.*, vol. 147, pp. 65–79, Jan. 2019, doi: 10.1016/j.isprsjprs.2018.10.017.

[36] M. Yang, C. Guo, H. Zhong, and H. Yin, "A curvature-based saliency method for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 9, pp. 1590–1594, Sep. 2021, doi: 10.1109/LGRS.2020.3005197.

[37] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1195–1204.

[38] Y. Zhou, X. Jiang, Z. Chen, L. Chen, and X. Liu, "A semisupervised arbitrary-oriented SAR ship detection network based on interference consistency learning and pseudolabel calibration," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 5893–5904, 2023, doi: 10.1109/JSTARS.2023.3284667.

[39] Z. Tian, W. Wang, K. Zhou, X. Song, Y. Shen, and S. Liu, "Weighted pseudo-labels and bounding boxes for semisupervised SAR target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 5193–5203, 2024, doi: 10.1109/JSTARS.2024.3363491.

[40] J. Shen, C. Zhang, Y. Yuan, and Q. Wang, "Enhancing prospective consistency for semisupervised object detection in remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5619312, doi: 10.1109/TGRS.2023.3310026.

[41] J. Wang, K. Chen, R. Xu, Z. Liu, C. C. Loy, and D. Lin, "CARAFE: Content-aware reassembly of features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 3007–3016, doi: 10.1109/ICCV.2019.00310.

[42] W. Shi et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1874–1883, doi: 10.1109/CVPR.2016.207.

[43] 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.

[44] T. Zhang et al., "Balance learning for ship detection from synthetic aperture radar remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 182, pp. 190–207, Dec. 2021, doi: 10.1016/j.isprsjprs.2021.10.010.

[45] T. Yue, Y. Zhang, J. Wang, Y. Xu, P. Liu, and C. Yu, "A precise oriented ship detector in SAR images based on dynamic rotated positive sample mining," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 10022–10035, 2023, doi: 10.1109/JSTARS.2023.3326163.

[46] Y. Liu, G. Yan, F. Ma, Y. Zhou, and F. Zhang, "SAR ship detection based on explainable evidence learning under intraclass imbalance," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5207715, doi: 10.1109/TGRS.2024.3373668.

[47] W. Wang, D. Han, C. Chen, and Z. Wu, "FastPFM: A multi-scale ship detection algorithm for complex scenes based on SAR images," *Connection Sci.*, vol. 36, no. 1, Dec. 2024, Art. no. 2313854, doi: 10.1080/09540091.2024.2313854.

[48] 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.

[49] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. SAR Big Data Era, Models, Methods Appl.*, 2017, pp. 1–6, doi: 10.1109/BIGSARDATA.2017.8124934.

[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, Mar. 2019, Art. no. 765, doi: 10.3390/rs11070765.

[51] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755, doi: 10.1007/978-3-319-10602-1_48.

**Man Chen** received the B.E. degree in electrical engineering and automation from Anhui Jianzhu University, Hefei, China, in 2019 and the M.S. degree in control engineering from the Changsha University of Science and Technology, Changsha, China, in 2022. He is currently working toward the Ph.D. degree in computer science and technology with the Army Engineering University of PLA, Nanjing, China.

His main research interests include image processing, remote sensing, and machine learning.

**Yuanlin He** received the M.S. degree in artificial intelligence from the Sichuan University of Science and Engineering, Yibin, China, in 2023. He is currently working toward the Ph.D. degree in computer science and technology with the Army Engineering University of PLA, Nanjing, China.

His main research interests include pattern recognition, image processing, and multimodal fusion.

**Tianfeng Wang** received the M.S. and Ph.D. degrees in computer science and technology from the Army Engineering University of PLA, Nanjing, China, in 2020 and 2023, respectively.

He is currently a Lecturer with the Army Engineering University of PLA. His main research interests include image processing, pattern recognition, and graph theory.

**Yahao Hu** received the B.S. degree in computer science and technology in 2019, from the Army Engineering University of PLA, Nanjing, China, where he is currently working toward the Ph.D. degree in computer science and technology.

His main research interests include natural language processing, multimodal fusion, and machine learning.

**Jun Chen** received the Ph.D. degree in cyberspace security from the Army Engineering University of PLA, Nanjing, China, in 2022.

He is currently a Lecturer with the Army Engineering University of PLA. His main research interests include remote sensing, image processing, and machine learning.

**Zhisong Pan** received the Ph.D. degree in computer science and technology from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2003.

He is currently a Professor with the Army Engineering University of PLA, Nanjing. His main research interests include deep learning, machine learning, and pattern recognition.