A Semisupervised Arbitrary-Oriented SAR Ship Detection Network Based on Interference Consistency Learning and Pseudolabel Calibration

Yue Zhou , Xue Jiang , *Senior Member, IEEE*, Zeming Chen, Lin Chen , and Xingzhao Liu , *Member, IEEE*

Abstract—The rapid development of deep learning cannot be achieved without the support of abundant labeled data. However, obtaining such a large amount of annotated data needs the support of professionals in the field of synthetic aperture radar (SAR) image understanding, which leads to the scarcity of SAR datasets with annotations. The scarcity of annotations poses a bottleneck in the performance of SAR ship detectors based on deep learning. Recently, semisupervised learning has become a hot paradigm, which can mine effective information from unlabeled data to further improve the performance of SAR ship detectors. However, existing semisupervised SAR ship detection studies all adopted multistage semisupervised frameworks, which are complex and inefficient. In this article, we first design an end-to-end semisupervised framework for SAR ship detection. To overcome the strong interferences resulting from the imaging or quantization processes in SAR, we introduce the interference consistency learning mechanism to enhance the model's robustness. To solve the complex background in the inshore scenario, a pseudolabel calibration network is designed to calibrate the pseudolabel according to the context knowledge around the ships. Based on the high-resolution SAR images dataset (HRSID) and the other four datasets, the superiority of the proposed approach over several state-of-the-art semisupervised frameworks has been evaluated under various labeling ratios, i.e., 1%, 5%, 10%, and 100%.

Index Terms—Arbitrary-oriented synthetic aperture radar (SAR) HRSID, ship detection, semisupervision.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) [1] ship detection plays a pivotal role in interpreting SAR images, which has been widely applied in maritime traffic safety and battlefield reconnaissance. Traditional SAR ship detectors can be divided into the following four categories:

1) threshold-based approaches [2], [3];
2) saliency-based approaches [4], [5];
3) hand-crafted feature-based approaches [6], [7];
4) statistical modeling-based approaches [8], [9].

With the emergence of deep learning, convolutional neural networks (CNNs) [10] and graph convolutional networks [11] have been applied in remote sensing and have shown great advantages against handcrafted features. These deep learning detectors outperformed the traditional methods in SAR ship detection tasks [12], [13], [14], [15]. However, these are horizontal bounding box (HBB) based detectors. Ships have arbitrary orientations and elongated shapes. HBB is not an effective representation of ships, particularly when ships are densely arranged near the wharf. To solve this problem, the oriented bounding box (OBB) is used to locate the target in remote sensing image (RSI) [16], [17], [18], [19], [20], [21]. At the same time, more and more high-resolution SAR datasets with OBB annotations have emerged [22], [23], [24], [25], which has led to significant progress in deep learning based SAR ship detectors. Cui et al. [26] and Su et al. [27] adapt the novel attention mechanism and deformable convolution to enhance the model's ability to extract key information from complex backgrounds. He et al. [28] designed a polar encoding to solve the problem of angle discontinuity in oriented SAT ship detection. Zhou et al. [29] attempts to replace CNN with transformer to introduce the global attention mechanism, which can further improve the detection performance of SAR ships in complex scenarios.

The above-mentioned models that must be trained with labeled data belong to supervised learning (SL). The dependence on the labeled data becomes the bottleneck of the existing CNN-based SAR ship detectors in real applications. On one hand, the data labeling of SAR is difficult. It can only be done with the support of professionals with the corresponding backgrounds. Researchers have attempted to use optical, infrared, and other sensor data to assist in improving the performance of models on SAR images [30], [31]. However, this requires paired annotated data, which is difficult to meet in some practical applications. On the other hand, massive SAR images are produced every day. These unlabeled data account for a much larger proportion than labeled data. Without labeling the targets in time, these massive SAR images cannot be used effectively. If considering the practical applications of the SAR ship detection tasks, the unlabeled data containing useful information cannot be neglected.

Recently, semisupervised learning (SSL) has become a hot paradigm, which can mine effective information from unlabeled

Yue Zhou, Xue Jiang, Lin Chen, and Xingzhao Liu are with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: sjtu_zy@sjtu.edu.cn; xuejiang@sjtu.edu.cn; linchenee@sjtu.edu.cn; xzhliu@sjtu.edu.cn).

Zeming Chen is with the Shenzhen International Graduate School, Tsinghua University, Shenzhen 518000, China (e-mail: czm20@mails.tsinghua.edu.cn).
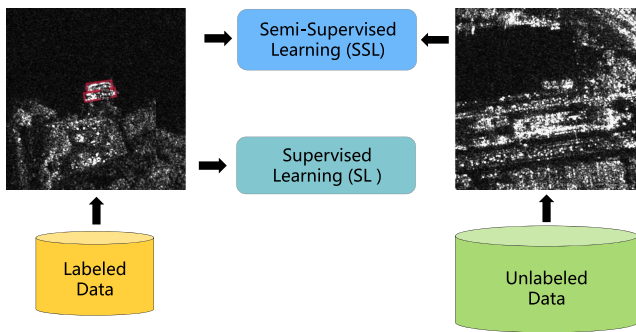
Fig. 1. Compared with SL, which only uses labeled data, SSL uses unlabeled and labeled data to train the network jointly. Since the high labeling cost and the existence of a massive number of unlabeled images, SSL is very suitable for SAR images.

data to further improve performance. Fig. 1 shows the difference between SL and SSL. Compared with SL, the training process of SSL involves unlabeled data, which means that the SSL paradigm has potential application value in some actual scenarios where massive unlabeled SAR image data can be obtained. Wang et al. [34] directly tried to apply the SSL based image classification to a SAR ship detection task. Chen et al. [35] designed a cross-domain coattention feature correlation module that addresses the domain adaptation problem. Hou et al. [36] designed an adversarial network to make the local features of the SAR ship in the unlabeled images closer to those in the marked images. However, existing SSL applications [32], [33] in SAR ship detection all adopt the multistage framework, which limits the training efficiency and performance. On one hand, the training process of the multistage framework is complex. As shown in Fig. 2, it includes three stages. It needs to train the teacher model first and then train the student model, so the training time is much longer than SL. On the other hand, the performance of the detector will be limited by the initial pseudolabels. Once pseudolabels of multistage frameworks are generated, they will not change throughout the entire training process. The errors in the initial pseudolabels will mislead the student model's learning.

Unlike the multistage methods, [37] introduced an end-to-end training strategy into an SSL, which has one training stage. During each iteration, the teacher model's parameters are updated by the student model's exponential moving average (EMA). As the pseudolabels of the end-to-end framework update dynamically, the quality of the generated pseudolabels will be improved gradually. However, various interferences may be introduced in SAR imaging and quantization, e.g., speckle noise and scattering interference [38]. Once the ship in the unlabeled image has been contaminated, it is difficult to detect it. The number of pseudolabels will be significantly reduced, and the potential information in unlabeled data cannot be fully mined. The quality of pseudolabels will also decrease, which may mislead the model's learning. These interferences are unique to SAR images and have not been discussed in previous end-to-end semisupervised frameworks. Moreover, SAR ships in inshore scenes are subject to serious background interference, which can impact the accuracy of pseudolabels and lead to false alarms, such as wharf buildings with shapes similar to ships. To solve

these issues, we propose an end-to-end semisupervised network for arbitrary-oriented SAR ship detection based on interference consistency learning (ICL) and pseudolabel calibration network (PLC). ICL enhances the model's robustness to different interferences in SAR images, thereby improving the quality of pseudolabels in interference scenarios. PLC is designed to calibrate the pseudolabel according to the context knowledge surrounding ships. It can alleviate the error caused by inaccurate pseudolabels in inshore scenarios, further improving detection accuracy. Previous SAR ship semisupervised work has not attempted to use contextual information to improve the quality of pseudolabels in inshore scenarios.

The main contributions of this work are summarized as follows.

1) We propose the first end-to-end semisupervised framework for SAR ship detection, which is more efficient than existing multistage semisupervised frameworks.
2) The ICL is introduced to enhance the model's robustness under strong interferences resulting from the imaging or quantization processes in SAR.
3) The PLC is introduced to calibrate the incorrect pseudolabel introduced by complex inshore backgrounds.
4) The experimental outcomes across high-resolution SAR images dataset (HRSID) and four other datasets demonstrate that our proposed approach outperforms several prevalent frameworks.

The rest of this article is organized as follows. The related works are introduced in Section II. In Section III, the SAR-Teacher is introduced, whose experimental results compared with other state-of-the-art methods are given in Section IV. Finally, Section V concludes this article.

Notations: Throughout the article, matrixes, vectors, and scalars are represented by bold uppercase letters X, bold lowercase letters x, and regular letters x, respectively. The primary notations used in this article are listed as follows.

Symbol	Definition
$\mathcal{L}_s$	The supervised loss
$\mathcal{L}_u$	The unsupervised loss
$\mathcal{L}_{\text{cls}}$	The classification loss
$\mathcal{L}_{\text{loc}}$	The location loss
$\mathcal{L}_p$	The supervised loss of PLC
$\mathbf{X}$	An unlabeled SAR image
$\mathcal{B}$	The set of rotated box (x, y, w, h, θ)
N_l	The number of proposals from labeled image
N_u	The number of proposals from unlabeled image
$\phi(\cdot)$	The teacher model
$\varphi(\cdot)$	The student model
$\mathcal{S}(\cdot)$	The operation to subject simulated interference on SAR image
$F(\cdot)$	The confidence threshold filter
σ	The interference consistency coefficient
IoU	The intersection over union between two rotated boxes

II. METHODOLOGY

A. Multistage Versus End-to-End Frameworks

This section compares the mainstream paradigms of multistage and end-to-end semisupervised frameworks for object detection, which are shown in Fig. 2.
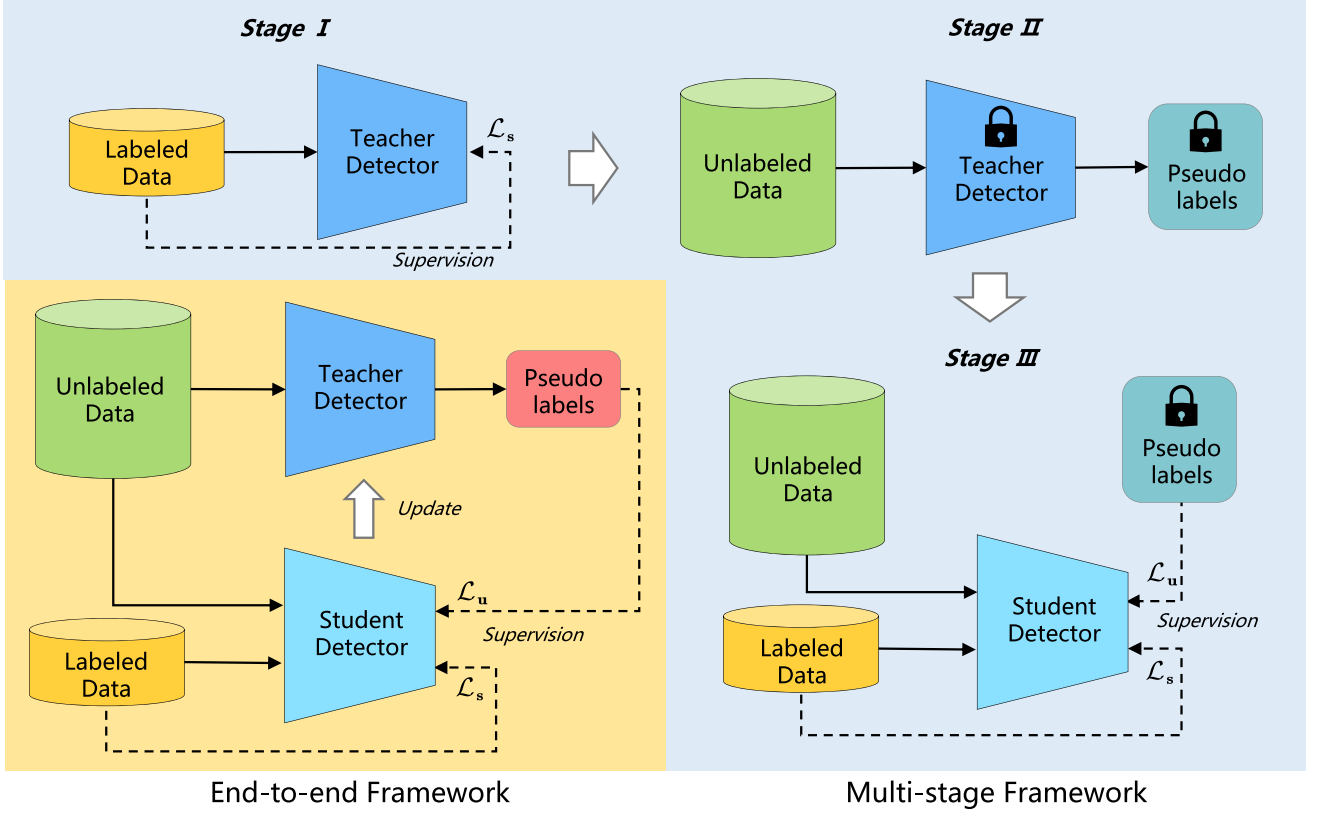
Fig. 2. Comparision of multi-tage and end-to-end semisupervised framework for object detection.

The multistage semisupervised object detection framework consists of the following three stages [37]:

1) In stage I, the teacher detector utilizes the labeled data for supervised training, whose loss is defined as follows:

$$\mathcal{L}_s = \frac{1}{N_l} \sum_{i=1}^{N_l} \left(\mathcal{L}_{cls}\left(\hat{t}_i, t_i\right) + t_i \sum_{j \in \mathcal{B}} \mathcal{L}_{\text{loc}}\left(\hat{v}_{ij}, v_{ij}\right) \right) \tag{1}$$

where j belongs to the set of (x, y, w, h, θ). The center coordinates, width, height, and angle of a rotated bounding box are represented by $x, y, w, h,$ and θ, respectively. N_l represents the number of proposals from labeled images. t_i is a binary value ($t_i = 1$ for ship, $t_i = 0$ for background), $\hat{t}_i$ denote the predicted probability value of ship class. v_{ij} and $\hat{v}_{ij}$ represent the ground truth and the predicted value of the ith rotated bounding box offsets. The location and classification loss are defined as follows:

$$\mathcal{L}_{\text{cls}} = -t_i \cdot \log\left(\hat{t}_i\right) - (1 - t_i) \cdot \log\left(1 - \hat{t}_i\right) \tag{2}$$

$$\mathcal{L}_{\text{loc}} = \begin{cases} 0.5\left(\hat{v}_{ij} - v_{ij}\right)^2 & |\hat{v}_{ij} - v_{ij}| < 1 \\ |\hat{v}_{ij} - v_{ij}| - 0.5 & \text{otherwise.} \end{cases} \tag{3}$$

2) In stage II, there is no training process. The teacher detector uses the model weights obtained in stage I to infer unlabeled images and generate pseudolabels. Usually, a detection score threshold is manually set here to ensure the pseudolabel's high quality. For multistage semisupervised

frameworks, once the pseudolabels are generated, they no longer change.

3) In stage III, the unlabeled images, with pseudolabels, and the labeled images, with grounded truth, work together in training the student model. The overall loss of stage III can be formulated as

$$\mathcal{L} = \mathcal{L}_s + \mathcal{L}_u \tag{4}$$

where $\mathcal{L}_s$ represents the supervised loss of labeled images, which is the same as stage I given in (1). $\mathcal{L}_u$ represents the unsupervised loss of the unlabeled images, which is defined as

$$\mathcal{L}_u = \frac{1}{N_u} \sum_{k=1}^{N_u} \left(\mathcal{L}_{cls}\left(\hat{t}_k, t_k^p\right) + p_k \sum_{j \in \mathcal{B}} \mathcal{L}_{\text{loc}}\left(\hat{v}_{kj}, v_{kj}^p\right) \right) \tag{5}$$

where N_u represents the number of proposals from unlabeled images. The superscript p indicates that the variable is a pseudolabel. The aim of most semisupervised frameworks is to design the unsupervised loss $\mathcal{L}_u$. However, the multistage semisupervised framework includes multiple training stages, and the training process is redundant and inefficient.

End-to-end frameworks have the same overall loss as the stage III of multistage frameworks, but the main difference lies in the joint optimization of student and teacher models [37]. Fig. 2 illustrates that the training process for end-to-end frameworks
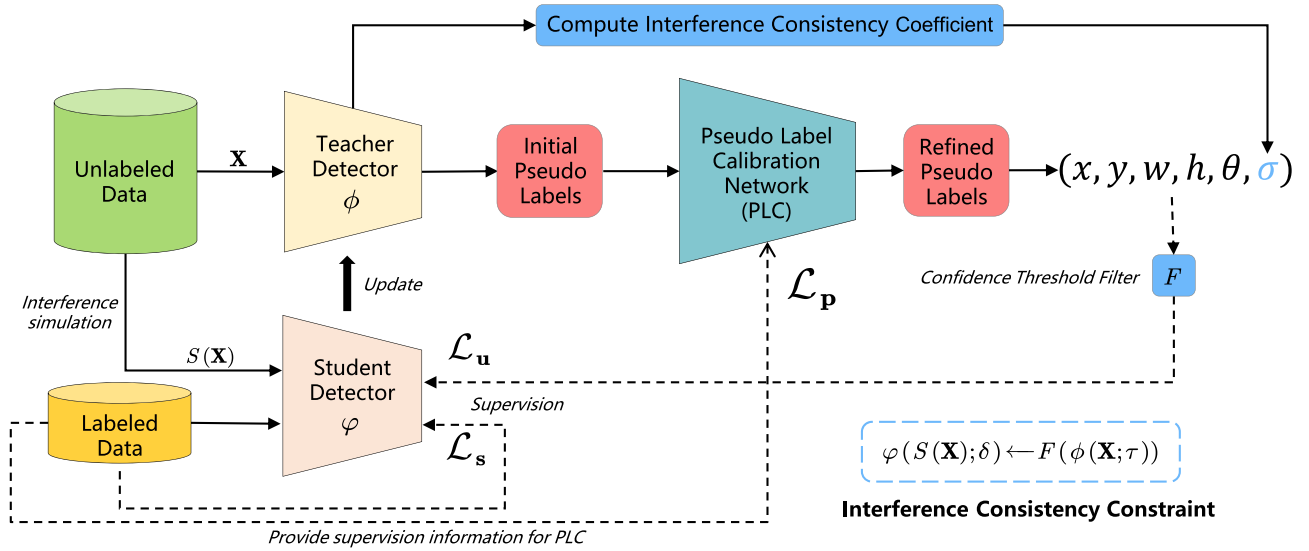
Fig. 3.   Overall architectures of the proposed semisupervised framework.

involves a single stage. In each iteration, the training data batch consists of the following two parts: 1) labeled and 2) unlabeled images half and half. The student model is then trained using both the labeled and unlabeled images. Upon completion of each iteration, the teacher model is updated through the EMA of the student model. This joint optimization strategy has the potential to improve the pseudolabel's quality dynamically, which is significant for SSL.

To our best knowledge, there are no existing works designing SAR ship detectors based on the end-to-end semisupervised framework for SAR ship detection. Due to the scattering interference and speckle noise in SAR images, the credibility of pseudolabels will be reduced. Incorrect pseudolabels will mislead the student model, damaging the ship detection performance. Thus, directly applying the end-to-end semisupervised frameworks designed for optical images cannot provide satisfactory results. We demonstrate the proposed end-to-end semisupervised framework for arbitrary-oriented SAR ship detection. The core is to improve the pseudolabel's quality for the application in SAR, including interference consistency learning and pseudolabel calibration, which is shown in Fig. 3.

B. Interference Consistency Learning

Previous SAR ship semisupervised work did not discuss the impact of various interference in SAR images on pseudolabels. However, in practical applications, these interferences cannot be ignored. They will increase the feature differences between the targets, leading to the model being prone to missing interfered targets. To suppress various interferences introduced in the imaging or quantization process of SAR images, we introduce interference consistency learning. First, we construct the interference consistency constraint by conducting interference simulations on SAR images. Then, we gauge the model's confidence about the pseudolabels by the interference consistency coefficient. We will introduce these two parts one by one.

1) Interference Consistency Constraint: The end-to-end semisupervised framework of this article adopts the teacher–student paradigm. We first feed the original unlabeled SAR images into the teacher model to generate pseudolabels. Then feed the same SAR image after simulated interference into the student model. Since the simulated interference does not change the position of the target, the pseudolabels generated by the teacher model can be directly used to provide supervision information for the student model. In this way, the framework can use unlabeled SAR images to construct interference consistency constraints.

For convenience, we mark the teacher model as ϕ and the student model as φ. Given an image $\mathbf{X}$, the teacher model output the results as follows:

$$\phi(\mathbf{X};\tau) = \{\hat{\mathbf{v}}_d\}_{d=1}^{D} \tag{6}$$

where τ represents the trainable parameters of the teacher model. D denotes the number of detection results. $\hat{\mathbf{v}}_d$ represents the dth predicted rotated bounding box vector. After passing through the filter F (e.g., a fixed confidence threshold), $\hat{\mathbf{v}}_d$ will become the pseudolabel $\mathbf{v}_d^{pre}$. Then, we implement the simulation interference on the same image to obtain the $\mathcal{S}(\mathbf{X})$, and construct a constraint as follows:

$$\varphi(\mathcal{S}(\mathbf{X});\delta) \xleftarrow{\text{supervision}} F(\phi(\mathbf{X};\tau)) \tag{7}$$

where δ represents the trainable parameters of the student model. $\mathcal{S}(\cdot)$ represents the operation to subject simulated interference on unlabeled SAR images, which are composed of seven different interference simulations as shown in Fig. 4:

1) speckle noise;
2) solarization;
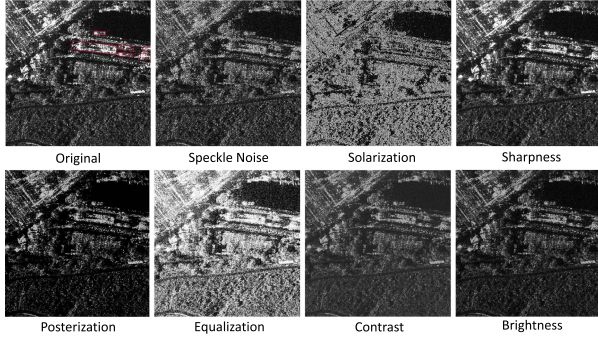3) sharpness;
4) posterization;
5) equalization;
6) contrast;

Fig. 4.   Interference simulations of SAR images.

Fig. 5.   Pseudo label calibration network.

7) brightness.

We train the student model under the simulation interference $S(\mathbf{X})$ to promote the student to learn the consistency of SAR image interference. This consistency knowledge will be transmitted to the teacher model as it is updated, thereby improving the quality of pseudolabels in unlabeled SAR images with speckle noise or scattering interference.

2) Interference Consistency Coefficient: In the previous methods, unsupervised loss treats all pseudolabels equally. However, there is a possibility of errors in pseudolabels. We need to pay more attention to reliable pseudolabels. Therefore, we introduce an interference consistency coefficient σ to measure the credibility of pseudolabels under severe interference, which ranges from 0 to 1. Each pseudolabel has a corresponding interference consistency coefficient. The larger the coefficient, the more confident the teacher model is in the current pseudolabel. On the contrary, smaller coefficients indicate a higher probability of the pseudolabel being incorrect.

Deep learning-based detectors allocate multiple proposals for each ground truth box. If the location of a target is uncertain, the model may generate multiple scattered proposals near the ground truth box. Conversely, if the model is certain about the location of a target, the generated proposals will be concentrated around the ground truth box. We aim to gauge the model's confidence about the pseudolabels by the proposals as follows:

$$\sigma^k = \frac{\sum_{i=1}^{N_p} \text{IoU}_{ik}}{N_p} \tag{8}$$

where $\sigma \in [0, 1)$. N_p denotes the number of positive samples assigned to the kth pseudolabel. IoU_{kj} denotes the rotated IoU between the ith positive sample and the kth pseudolabel. IoU is defined as the overlap area divided by the union area of two rotated boxes.

Next, σ is used to weight the unsupervised loss of the unlabeled images $\mathcal{L}_u$. The larger the σ, the greater the $\mathcal{L}_u$, and vice versa. Thus, the σ can reduce the attention of the model to uncertain pseudolabels, and put more attention on pseudolabels with higher reliability.

C. Pseudolabel Calibration Network

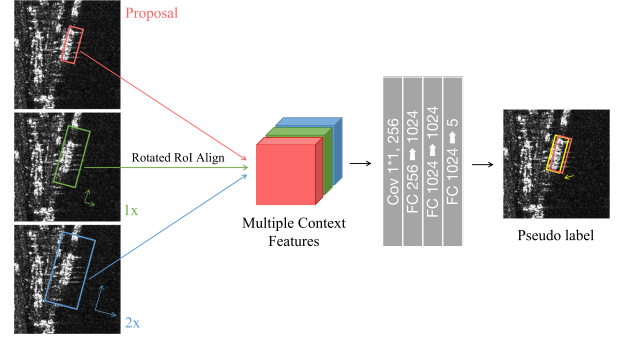Another challenge in SAR ship detection is the presence of numerous small ships overwhelmed by complex surrounding backgrounds in inshore scenes. The complex background can blur targets' boundaries, leading to inaccurate positioning and reducing pseudolabel's quality, hampering the student model's performance. It has been proven that exploiting the context knowledge around small objects is beneficial to locate their bounding box accurately [46]. However, previous SAR ship semisupervised works did not attempt to use contextual information to improve the quality of pseudolabels in inshore scenarios. Motivated by this idea, we have designed a pseudolabel calibration network, which can calibrate the pseudolabels based on the context knowledge around the ship and improve the quality of the pseudolabel.

The structure of the pseudolabel calibration network is shown in Fig. 5. First, we enlarge each proposal with factors $\{2, 4\}$, and get two additional proposals. Second, we feed these three proposals into the rotated region of interest (RoI) Align operator [45] to obtain multiple contextual features, which contain information from the target and surrounding background. Then, we concatenate the three features and feed them into a lightweight information fusion module. This module consists of a convolutional layer and three fully connected layers. By calculation, the computational complexity of the lightweight information fusion module is 20.875 GFlops. Finally, we can obtain the pseudolabels after calibration.

We use the labeled SAR images to train the PLC network in a supervised method. The PLC loss $\mathcal{L}_p$ can be formulated as

$$\mathcal{L}_p = \frac{1}{N_l} \sum_{i=1}^{N_l} \left(1 - \text{IoU}_i^2\right) \tag{9}$$

where IoU_i denotes the rotated IoU between the ith positive sample and the ground truth assigned to it. It should be noted that PLC loss only applies to the training of labeled images. Then, we update its weight in the student model to the teacher model for calibrating pseudolabels.

PLC can appropriately mitigate the deviation between the pseudo labels and the ground truth and improve the quality of pseudolabels. It is worth mentioning that PLC is only used during training phase and can be removed during testing phase. Therefore, PLC does not increase the computational complexity and inference speed of the testing phase.
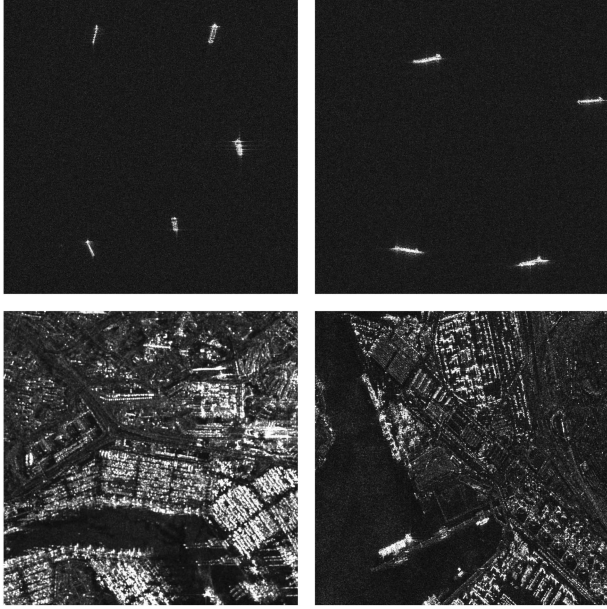
Fig. 6. Example of an HRSID dataset.

D. Loss Function

Finally, the loss of the proposed end-to-end framework can be formulated as

$$
\begin{aligned}
\mathcal{L} &= \mathcal{L}_s + \mathcal{L}_u + \mathcal{L}_p \\
&= \frac{1}{N_l} \sum_{i=1}^{N_l} \left(\mathcal{L}_{\text{cls}} \left(\hat{t}_i, t_i \right) + t_i \sum_{j \in \mathcal{B}} \mathcal{L}_{\text{loc}} \left(\hat{v}_{ij}, v_{ij} \right) \right) \\
&\quad + \frac{1}{N_u} \sum_{k=1}^{N_u} \sigma_k \left(\mathcal{L}_{\text{cls}} \left(\hat{t}_k, t_k^p \right) + p_k \sum_{j \in \mathcal{B}} \mathcal{L}_{\text{loc}} \left(\mathcal{S}(\hat{v}_{kj}), v_{kj}^p \right) \right) \\
&\quad + \frac{1}{N_l} \sum_{i=1}^{N_l} \left(1 - \text{IoU}_i^2 \right)
\end{aligned}
\tag{10}
$$

where σ^k denotes interference consistency coefficient of the kth pseudolabel, which reduces the possibility of student models being misled by incorrect pseudolabels. $\mathcal{S}(\hat{v}_{kj})$ are the predicted offsets of the ith rotated bounding box of simulated interference SAR images. Since the proposed simulated interferences are invariant to the box coordinates, the operation $\mathcal{S}$ does not need to apply on the v_{kj}^p.

The proposed end-to-end framework uses the EMA strategy to update the teacher model after each iteration. This joint optimization strategy can dynamically improve pseudolabel's quality.

III. EXPERIMENT AND ANALYSIS

A. Dataset Description and Implementation Details

Experiments are performed on the HRSID [23], which is a ship detection dataset from high-resolution SAR images with massive ship samples. As can be seen from Fig. 6, it includes both offshore scenes with clean backgrounds and inshore scenes

with complex backgrounds. It also has various scattering interferences and speckle noise. In addition, the other four SAR ship datasets are also used as unlabeled in the extended experiments. It contains three OBB datasets [22], [24], [25] and one HBB dataset [44]. It should be noted that we did not use the annotation of these four datasets during the whole experiment. The detailed information of these open-source SAR ship datasets is shown in Table I.

Our algorithm implementation and hyper-parameters settings are based on a unified rotated object detection tool-box (MMRotate) [45]. For all datasets, the image shape for network input is 800×800 pixels. Our experiments were conducted on a CentOS 7.3 system with an RTX 3090Ti graphics processing unit (GPU). The variance of speckle noise is set to 1. We referred to the [33] and used the following two experimental settings in this article: 1) partially labeled data and 2) fully labeled data. We will introduce the setting details of them below, which has significant differences.

In the partially labeled data setup, we randomly sample 1%, 2%, 5%, and 10% images from the HRSID training set as labeled training data, and the images that are not selected are considered as the unlabeled data. The overall number of HRSID training set images is 3623, which means that 1% of the training set has only 36 images. Therefore, this experimental setup can verify the few-shot learning ability of the model. To guarantee the experiment's rigor, we randomly choose five different data folds for each data proportion, and the final performance is calculated as the mean of the five folds. The training process involves running each model for 6 k iterations on one GPU with stochastic gradient descent (SGD) as the optimizer training. The learning rate equals to 0.0025 and decreased by a factor of 10 after 4 k and 5.5 k iterations. Each batch comprises of four images, which includes two labeled and two unlabeled samples. Moreover, weight decay and momentum are set at 0.0001 and 0.9, respectively, for all models.

In the fully labeled data setup, the fully HRSID training set is utilized as labeled data, and supplement it with unlabeled data from other SAR ship datasets. The goal is to verify the performance of semisupervised models on large-scale unlabeled samples. It is more consistent with real-world application scenarios and more challenging. Each model is trained for 65 k iterations on one GPU. The learning rate equals to 0.0025 and decreased by a factor of 10 after 43 k and 60 k iterations.

B. Evaluation Metrics

We use average precision (AP) to quantitatively evaluate the detection performance and the latency per image to evaluate inference speed

$$
P = \frac{N_{tp}}{N_{\text{pred}}},
\tag{11}
$$

$$
R = \frac{N_{tp}}{N_{\text{targets}}},
\tag{12}
$$

$$
AP = \int_0^1 P(R) \, dR
\tag{13}
$$

TABLE I
DETAILED INFORMATION OF THE OPEN-SOURCE SAR SHIP DATASETS USED IN THIS ARTICLE

Dataset	Year	Sources	Resolution(m)	Image Size	# of Images	# of Ships	Classes	Annotations
SSDD	2017	RadarSat-2 TerraSAR-X Sentinel-1	$1\sim15$	$214\sim668$	1160	2540	1	OBB
HRSID	2020	Sentinel-1 TerraSAR-X	0.5,1,3	800	5604	16951	1	OBB
LS-SSDD-v1.0	2021	GF-3	5×20	24000×16000	15	6016	1	HBB
SRSDD-v1.0	2021	GF-3	1	1024	666	2884	6	OBB
RSDD-SAR	2022	TerraSAR-X GF-3	$2\sim20$	512	7000	10263	1	OBB

TABLE II
DEFINITION OF THE METRICS USED IN THE EXPERIMENTS

Metric	Definition
Training time (s) [13]	Time cost of model training.
Training weight (MB) [14]	Hard disk space occupied by the model training weight file.
Testing weight (MB) [14]	Hard disk space occupied by the model inference weight file.
Training memory (GB) [14]	GPU memory space occupied by the model during training.
Testing memory (GB) [14]	GPU memory space occupied by the model during inference.
Iteration time (s) [13]	Time cost for each iteration of a data batch during the training process.
FPS (Images/s) [13]	Number of SAR images detected by the model per second.

TABLE III
COMPARISON OF MULTISTAGE END-TO-END FRAMEWORKS ON HRSID (1% PARTIALLY LABELED DATA)

Method	AP_{50}	Training time
Supervised (RetinaNet-OBB)	32.5	0:23:26
Multi stage (RetinaNet-OBB)	32.7	0:52:23
E2E (RetinaNet-OBB)	**35.5**	0:28:39
Supervised (ORCNN)	54.3	0:40:27
Multi stage (ORCNN)	53.4	2:23:56
E2E (ORCNN)	**59.9**	1:10:09

The bold items denote the optimal values for RetinaNet-OBB and ORCNN.

where N_{pred} represents the total number of predicted boxes, N_{tp} is the number of targets correctly detected, N_{target} denotes the actual number of targets, $P(R)$ is precision-recall curve, and AP is mean AP.

To measure the efficiency performance of different models more comprehensively, we also adopt seven evaluation metrics in Table II.

C. Experiments and Analysis

1) Effect of End-to-End: We compare the end-2-end method with the multistage framework on 1% partially labeled HRISD as shown in Table III. When using RetinaNet OBB as the detector, the performance of the multistage semisupervised framework is only 0.2 points higher than that of the supervision framework, while the performance of the end-to-end framework is 3.0 points higher than that of the supervision framework. Next, we replace the detector with the oriented RCNN (ORCNN) and repeat the above experiment. The E2E method also outperforms the multistage framework and supervised framework. However, the performance of ORCNN's multistage framework is even

TABLE IV
ABLATION EXPERIMENT OF VARIOUS INTERFERENCES ON HRSID (1% PARTIALLY LABELED DATA)

Simulated interference transforms	Speckle noise	AP_{50}
		59.9
✓		62.5
	✓	62.2
✓	✓	**63.2**

The bold items denote the optimal value in the column.

TABLE V
EFFECT OF INTERFERENCE CONSISTENCY LEARNING ON INSHORE AND OFFSHORE SCENES (PARTIALLY LABELED DATA)

ICL	Scene	1%	2%	5%	10%
	Inshore	18.1	22.8	26.8	29.3
✓	Inshore	**21.5**	**26.2**	**29.6**	**32.1**
	Offshore	80.9	85.3	**87.6**	**88.5**
✓	Offshore	**82.5**	**86.0**	87.4	**88.5**

The bold items denote the optimal values for inshore and offshore scenes.

TABLE VI
ABLATION EXPERIMENT OF SAR-TEACHER ON HRSID (PARTIALLY LABELED DATA)

ICL	PLC	1%	2%	5%	10%
		59.9	65.0	67.7	71.0
	✓	61.0	66.4	68.2	71.8
✓		63.2	67.2	69.3	72.8
✓	✓	**63.9**	**67.9**	**69.8**	**73.5**

The bold items denote the optimal value in the columns.

worse than the supervised framework. Meanwhile, the train time of the end-to-end semisupervised framework is only half that of the multistage semisupervised framework, significantly improving training efficiency. Compared with SL, the training time of end-to-end semisupervised framework increased slightly. This experiment verifies that the end-to-end scheme is not only superior to the multistage scheme in performance but also more versatile for the different detectors. To make the results of the semisupervised algorithm more competitive, we use the ORCNN as the baseline detector in our following experiments.

2) Ablation Experiment: We conducted ablation experiments on different interference combinations, and the experimental results are shown in Table IV. The experimental results show that the speckle noise can be used with random color transformations for better performance. The quantitative comparison of ICL is shown in Table V. We can see that the AP_{50} of ICL is highest in both offshore and inshore scenes. The AP_{50} of 1%

TABLE VII
COMPARISON WITH STATE-OF-THE-ART METHODS (PARTIALLY LABELED DATA)

Detector	Method	Proportion				Training			Testing		
		1%	2%	5%	10%	Weight	Memory	Iteration time	Weight	Memory	FPS
RetinaNet [39]	Supervised	32.5	47.4	54.3	54.9	140.0	**1.94**	**0.23**	140.0	0.14	**36.4**
FRCNN [40]	Supervised	47.1	52.0	60.2	69.3	159.1	2.30	0.36	159.1	0.16	8.0
ORCNN [41]	Supervised	54.3	61.9	64.6	70.9	159.1	2.30	0.37	159.1	0.16	6.6
LSKNet [42]	Supervised	60.7	64.9	66.9	68.8	**119.3**	4.22	0.49	**119.3**	**0.12**	7.7
ORCNN	MeanTeacher [43]	49.2	61.0	65.8	70.2	317.6	4.48	0.87	159.1	0.16	6.7
ORCNN	SoftTeacher [37]	54.8	62.9	66.5	70.8	317.6	4.21	0.90	159.1	0.16	6.3
LSKNet	Ours	**67.2**	**70.7**	69.1	70.1	346.3	9.12	1.08	**119.3**	0.17	9.6
ORCNN	Ours	63.9	67.9	**69.8**	**73.5**	**425.2**	5.36	1.00	159.1	0.21	9.8

The bold items denote the optimal value in the columns.

train set outperforms the baseline by 3.4 points in the inshore scenario. And ICL finally increases the AP_{50} of 1% train set by 1.6 points in offshore scenes. The experimental results show that ICL shows more vital ability in complex backgrounds. The ablation experiment of ICL and PLC is shown in Table VI. The experimental results show that PCL can be used with ICL to enhance the performance of our semisupervised framework further.

3) Comparison with Representative Methods: This part compares SAR-Teacher with several state-of-the-art end-to-end semisupervised frameworks for object detection. The quantitative comparison between the proposed framework and other frameworks on the partially labeled data is given in Table VII. The AP of the proposed framework is the highest in 1%, 2%, 5%, and 10%. We can also find that the two-stage detector, Faster RCNN (FRCNN) and ORCNN, has better learning ability than the single-stage detector (RetinaNet) when the train data are less. MeanTeacher performs poorly when the proportion of labeled data is low. SoftTeacher's performance on the SAR dataset is subpar, which highlights the significant gap between SAR and general optical images. Therefore, it is essential to devise a semisupervised framework specifically for SAR images. Moreover, we add the large selective kernel network (LSKNet) [42] detector to validate the effectiveness of the proposed method. Compared to ORCNN, LSKNet has a significant improvement effect when the proportion of labeled data is low, while its improvement effect is not as good as ORCNN when the proportion of labeled data is higher. Therefore, in the fully labeled data experiments, we chose ORCNN as our basic detector.

Besides AP under different proportions of labeled data, we also used six metrics to measure the efficiency performance of different models more comprehensively. Among them, train memory and iteration time belong to the training phase, while the other three metrics belong to the inference phase. According to Table VII, although the proposed semisupervised framework has higher spatial complexity, its time complexity is similar to other semisupervised methods.

Although the semisupervised object detection framework based on the student–teacher architecture has two object detectors during the training phase, only the student detectors need to be retained during the testing phase. Therefore, we used

two sets of indicators to separately evaluate the computational complexity and speed of the model during the training and testing phases. Although our proposed framework will increase the model complexity of the training process, it will not affect the model's inference speed and memory usage during the testing phase.

We also compare the qualitative results of the proposed method to those of the supervised baseline under 1% HRSID train set are shown in Fig. 7. By comparing the yellow and green ellipses in the figure, it can be found that SAR-Teacher has detected many targets that are missed in the supervised baseline.

4) Validation on Full Labeled Data: In our previous experiment, a portion of the HRSID training set was employed as the labeled dataset, while the remaining images acted as unlabeled data. To verify the performance of this semisupervised framework on the full HRSID dataset, we conducted a comparative experiment on the other four unlabeled datasets, which are shown in Table VIII. When SSDD is used as the unlabeled dataset, the semisupervised framework is 1.4 points higher than the fully supervised model. The advantages of the semisupervised framework become more evident as the number of unlabeled samples increases. When three SAR datasets are used as unlabeled datasets, the semisupervised framework is 2.5 points higher than the fully supervised model. This shows that the semisupervised framework proposed in this article has the ability to mine information from massive unlabeled SAR images, which can reduce the time spent by researchers on labeling images.

5) Validate on Data Including Pure Background Images: The images in the above datasets are obtained after being cut and filtered by researchers. Each picture slice contains at least one target, virtually reducing learning difficulty. However, in the actual application scenario, the obtained images are usually large-scale SAR images, and we need to split them into slices. Because we do not have annotations, we cannot filter out slices without targets. To verify the performance of this semisupervised framework in real application scenarios, we conducted a comparative experiment on whether the unlabeled dataset contains pure background images. We selected the large-scale SAR ship detection dataset, LS-SSDD, as the unlabeled samples for the experiment. We split the original images into 800 * 800 slices
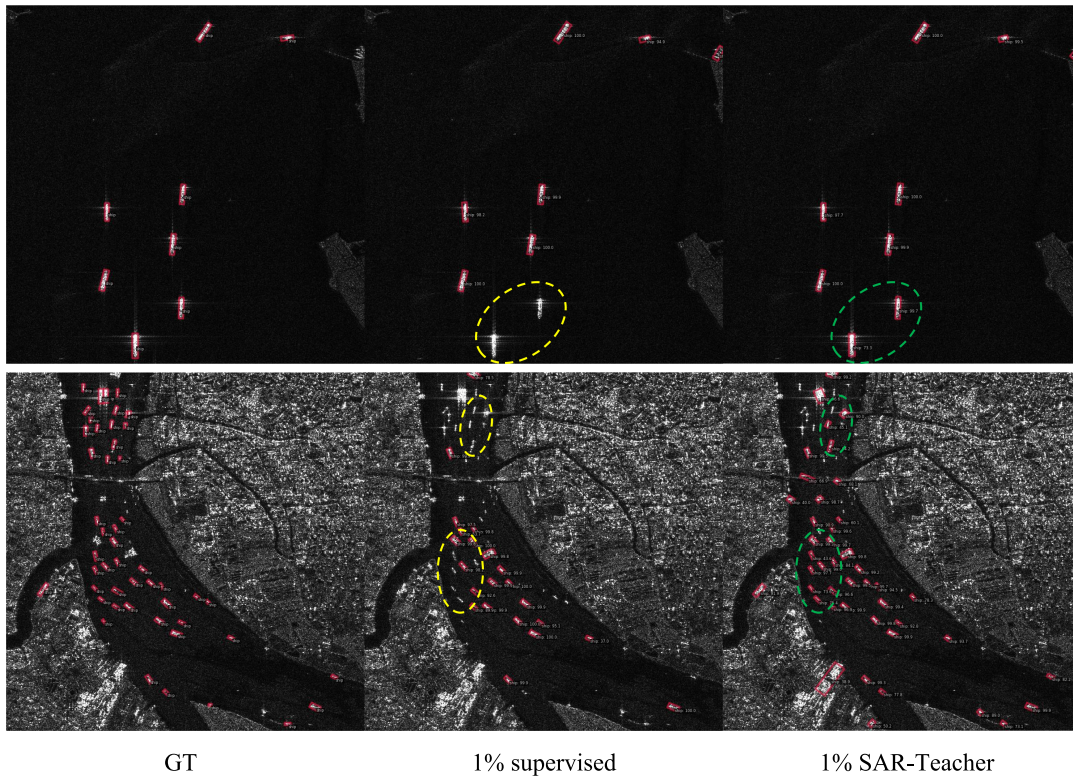
| GT | 1% supervised | 1% SAR-Teacher |

Fig. 7. Qualitative results of our semisupervised method on HRSID (1% partially labeled data).

TABLE VIII
VALIDATION ON FULLY LABELED DATA

Method	Labeled dataset	Unlabeled dataset	AP_{50}
Supervised	HRSID (100% training set)	-	80.7
SAR-Teacher	HRSID (100% training set)	SSDD	82.1
SAR-Teacher	HRSID (100% training set)	{SSDD, SRSDD}	83.0
Soft-Teacher	HRSID (100% training set)	{SSDD, SRSDD, RSDD}	81.4
SAR-Teacher	HRSID (100% training set)	{SSDD, SRSDD, RSDD}	**83.2**

The bold items denote the optimal value in the columns.

TABLE IX
COMPARATIVE EXPERIMENT ON WHETHER THE UNLABELED DATASET
CONTAINS PURE BACKGROUND IMAGES

Labeled dataset	Unlabeled dataset	AP_{50}
HRSID (100% training set)	LS-SSDD (w/o pure background)	81.3
HRSID (100% training set)	LS-SSDD	**82.4**

The bold items denote the optimal value in the column.

(try to keep consistent with the train set of HRSID). There are a total of 9000 slices, of which 1859 contain ships. The experimental results are shown in Table IX. The first experiment only used 1859 slices with ships, and the second used 9000 slices. The experiment proves that the pure background images do not weaken the training processing of the semisupervised framework but improve the detector's performance by 1.1 points. It is exciting because it also means that the semisupervised frame proposed in this article can be applied to SAR ship detection in real scenes. Unfortunately, due to the limited number of images

in the open-source SAR ship datasets, it is difficult to further improve the performance of our semisupervised algorithm.

6) Limitations of SAR-Teacher: We have conducted the full data experiments on the other four datasets separately, and the results are shown in Table X. The number of unlabeled samples does not play a decisive role in the results. Although SRSDD has the smallest number of images, it significantly improves the semisupervised model. On the contrary, the RSDD dataset, which has the largest number of images and ships, brings limited improvement to the semisupervised model. We believe that the resolution and image size play a decisive role. Only when the image of the unlabeled dataset is resized to make the ship size close to the labeled dataset, can the performance of the semisupervised framework be maximized. This shows that SAR-Teacher cannot accurately produce pseudolabels for unlabeled datasets with large differences in resolution. Therefore, we recommend that SAR images with the same resolution be used as unlabeled samples in the actual application of the SAR-Teacher. However, if there are no unlabeled samples

TABLE X
ANALYZE THE CHARACTERISTICS OF UNLABELED SAMPLES AND THEIR IMPACT ON THE RESULTS

Dataset	Type	Sources	Resolution(m)	Image Size	# of Images	# of Ships	AP_{50}
HRSID	Labeled	Sentinel-1 TerraSAR-X	0.5,1,3	800	5604	16951	-
SSDD	Unlabeled	RadarSat-2 TerraSAR-X Sentinel-1	1∼15	214∼668	1160	2540	82.1
SRSDD-v1.0	Unlabeled	GF-3	1	1024	666	2884	**82.9**
RSDD-SAR	Unlabeled	TerraSAR-X GF-3	2∼20	512	7000	10263	82.4

The bold items denote the optimal value in the column.

TABLE XI
EXPERIMENT SETTINGS OF KNOWLEDGE FORGETTING AND ERROR ACCUMULATION ON HRSID DATASET

Experiment	Labeled training set	Unlabeled training set	Testing set
Normal semi supervised	1% training set	99% training set	All testing set
Knowledge forgetting	1% training set	99% training set	1% training set
Error accumulation	1% training set	99% training set	99% training set

of the same resolution, we can also improve the model's performance by increasing the number of unlabeled samples. As shown in Table VIII, more unlabeled samples can also improve the performance of semisupervised models.

IV. DISCUSSION

A. Knowledge Forgetting and Error Accumulation Issues

As shown in Table XI, the experimental setup in this section is slightly different from the experimental section above. To explore the knowledge-forgetting problem of the model, we replaced the testing set with a 1% training set to see if the model had forgotten the initial supervised data. To examine the problem of error accumulation in the model, we replaced the testing set with 99% of the training set, which can show whether the performance of the model on unlabeled data will gradually decrease.

The problem of knowledge forgetting in deep learning models was first discussed in [47]. When a trained model on a task is trained on a new task, previous knowledge may be severely forgotten. To verify whether the semisupervised object detection framework will suffer from sample forgetting, we conduct an experiment as shown in Fig. 8. From the experimental results, both semisupervised frameworks improve detection accuracy compared to supervised models. For the multistage semisupervised framework, the introduction of pseudolabels not only does not cause knowledge forgetting but also improves the detection accuracy of labeled data. It indicates that the semisupervised detection framework does not have the problem of knowledge forgetting.

Error accumulation is another issue that pseudolabels may cause, which was also first discussed in [47]. The main manifestation is that as the training progresses, the detection performance of the model decreases due to being misled by false labels. For multistage detection frameworks, once pseudolabels are generated, they are invariant. According to Fig. 9, it can be seen that both semisupervised frameworks improve detection accuracy compared to supervised models. For the multistage semisupervised framework, the number of erroneous labels does gradually decrease as training progresses. It indicates that the
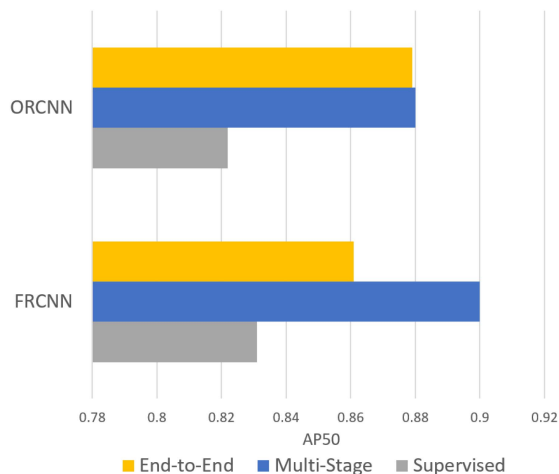

Fig. 8. Evaluation knowledge forgetting on 1% labeled data.

semisupervised detection framework does not have the problem of error accumulation.

V. CONCLUSION

In this article, an end-to-end semisupervised framework, SAR-Teacher, is proposed for arbitrary-oriented SAR ship detection. It can significantly reduce the detector's demand for labeled SAR images, which can break through the bottleneck of SL SAR ship detection. Specifically, we designed ICL to construct interference consistency constraints, which can prompt the learning ability of the model for ship detection from SAR images. Moreover, we propose a PLC network that utilizes contextual knowledge around the ship to calibrate incorrect pseudolabels and reduce the negative impact of complex inshore backgrounds. The superiority of the proposed semisupervised framework is verified through experiments conducted on only 1% labeled and fully labeled data.

Semisupervised technology is crucial for the intelligent interpretation of synthetic aperture radar ships. In the future, we will attempt to improve the efficiency of the semisupervised object detection framework and consider different application scenarios. On one hand, not all unlabeled images have a learning
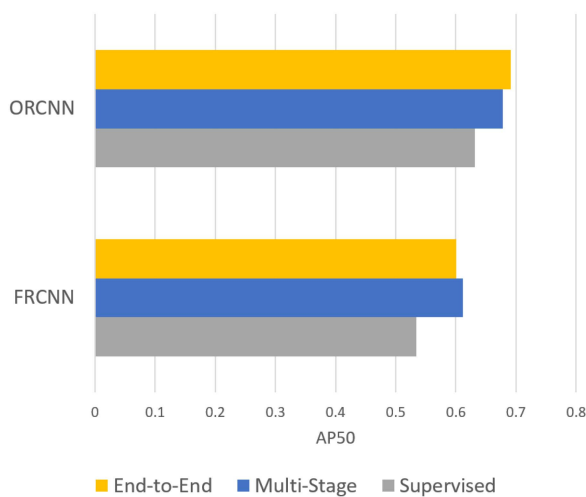
Fig. 9.   Evaluation error accumulation on 99% unlabeled data.

necessity, such as duplicate or similar images. Therefore, we can incorporate active learning techniques to select unlabeled images that are more worthy of learning and improve the training efficiency of the semisupervised target detection framework. On the other hand, this article only utilizes rotated box annotations. In fact, we can also utilize other SAR datasets with annotation difficulty lower than that of rotated box detection tasks, such as horizontal box detection tasks and scene classification tasks. The datasets for these tasks are easier to obtain online. In this way, the framework can fully utilize various labeled and unlabeled data, further expanding its application scenarios.

REFERENCES

[1] Y. Cui, G. Zhou, J. Yang, and Y. Yamaguchi, "On the iterative censoring for object detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 641–645, Jul. 2011.

[2] H. Dai, L. Du, Y. Wang, and Z. Wang, "A modified CFAR algorithm based on object proposals for ship object detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1925–1929, Dec. 2016.

[3] O. Pappas, A. Achim, 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.

[4] F. Gao, F. Ma, J. Wang, J. Sun, E. Yang, and H. Zhou, "Visual saliency modeling for river detection in high-resolution SAR imagery," *IEEE Access*, vol. 6, pp. 1000–1014, 2018.

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

[6] 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. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2376–2388, Jul. 2018.

[7] 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, Jan. 2018, Art. no. 190.

[8] G. Gao, K. Ouyang, Y. Luo, S. Liang, and S. Zhou, "Scheme of parameter estimation for generalized gamma distribution and its application to ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1812–1832, Mar. 2017.

[9] X. Wu, D. Hong, J. Tian, J. Chanussot, W. Li, and R. Tao, "ORSIm detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5146–5158, Jul. 2019.

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

[11] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[12] 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.

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

[14] 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.

[15] 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.

[16] Y. Long, X. Jiang, X. Liu, and Y. Zhang, "SAR ATR with rotated region based on convolution neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1184–1187.

[17] Y. Zhou, X. Jiang, Z. Li, and X. Liu, "Arbitrary-oriented SAR ship detection via frequency learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 4552–4555.

[18] J. Wang, C. Lu, and W. Jiang, "Simultaneous ship detection and orientation estimation in SAR images based on attention module and angle regression," *Sensors*, vol. 18, Aug. 2018, Art. no. 2856.

[19] S. Chen, J. Zhang, and R. Zhan, "R2FA-Det: Delving into high-quality rotatable boxes for ship detection in SAR images," *Remote Sens.*, vol. 12, no 12, Jun. 2020, Art. no. 2031.

[20] Z. Pan, R. Yang, and A. Zhang, "MSR2N: Multi-stage rotational region based network for arbitrary-oriented ship detection in SAR images," *Sensors*, vol. 20, no 8, Apr. 2020, Art. no. 2040.

[21] Q. An, Z. Pan, L. Liu, and H. You, "DRBox-v2: An improved detector with rotatable boxes for object detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8333–8349, Nov. 2019.

[22] T. Zhang, X. Zhang, J. Li, and X. Xu, "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sens.*, vol. 13, no 18, Sep. 2021, Art. no. 3690.

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

[24] S. Lei et al., "SRSDD-v1.0: A high-resolution SAR rotation ship detection dataset," *Remote Sens.*, vol. 13, no. 24, Dec. 2021, Art. no. 5104.

[25] X. U. Congan et al., "RSDD-SAR: Rotated ship detection dataset in SAR images," *J. Radars*, vol. 11, no. 4, pp. 581–599, 2022.

[26] Z. Cui, X. Wang, N. Liu, Z. Cao, and J. Yang, "Ship detection in large-scale SAR images via spatial shuffle-group enhance attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 379–391, Jan. 2021.

[27] N. Su, Z. Huang, Y. Yan, C. Zhao, and S. Zhou, "Detect larger at once: Large-area remote-sensing image arbitrary-oriented ship detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jan. 2022, Art. no. 6505605, doi: 10.1109/LGRS.2022.3144485.

[28] Y. He, F. Gao, J. Wang, A. Hussain, and H. Zhou, "Learning polar encodings for arbitrary-oriented ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3846–3859, 2021, doi: 10.1109/JSTARS.2021.3068530.

[29] Y. Zhou, X. Jiang, G. Xu, X. Yang, X. Liu, and Z. Li, "PVT-SAR: An arbitrarily oriented SAR ship detector with pyramid vision transformer," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 291–305, 2023, doi: 10.1109/JSTARS.2022.3221784.

[30] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[31] X. Wu, D. Hong, and J. Chanussot, "UIU-Net: U-Net in U-Net for infrared small object detection," *IEEE Trans. Image. Process.*, vol. 32, pp. 364–376, 2023, doi: 10.1109/TIP.2022.3228497.

[32] J. Jeong, S. Lee, J. Kim, and N. Kwak, "Consistency-based semi-supervised learning for object detection," in *Proc. Conf. Neural Inf. Process. Syst.*, 2019, pp. 10759–10768.

[33] 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*.

[34] C. Wang, J. Shi, Z. Zou, W. Wang, Y. Zhou, and X. Yang, "A semi-supervised SAR ship detection framework via label propagation and consistent augmentation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 4884–4887.

[35] S. Chen, R. Zhan, W. Wang, and J. Zhang, "Domain adaptation for semi-supervised ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4507405, doi: 10.1109/LGRS.2022.3171789.

[36] B. Hou et al., "A neural network based on consistency learning and adversarial learning for semisupervised synthetic aperture radar ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5220816, doi: 10.1109/TGRS.2022.3142017.

[37] M. Xu et al., "End-to-end semi-supervised object detection with soft teacher," in *Proc. Int. Conf. Comput. Vis.*, Montreal, QC, Canada, 2021, pp. 3040–3049.

[38] L. Chen, X. Jiang, Z. Li, X. Liu, and Z. Zhou, "Feature-enhanced speckle reduction via low-rank and space-angle continuity for circular SAR target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7734–7752, Nov. 2020.

[39] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[40] 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.

[41] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, "Oriented R-CNN for object detection," in *Proc. Int. Conf. Comput. Vis.*, 2021, pp. 3500–3509.

[42] Y. Li, Z. Zheng, M. Chen, J. Yang, and X. Li, "Large selective kernel network for remote sensing object detection," 2023, *arXiv:2303.09030*.

[43] I. Guyon et al., "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Conf. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1195–1204.

[44] T. Zhang et al., "LS-SSDD-v1.0: A deep learning dataset dedicated to small ship detection from large-scale sentinel-1 SAR images," *Remote Sens.*, vol. 12, no. 18, Sep. 2020, Art. no. 2997.

[45] Y. Zhou et al., "Mmrotate: A rotated object detection benchmark using pytorch," in *Proc. ACM Int. Conf. Multimedia*, 2022, pp. 7331–7334.

[46] X. Yang, J. Yang, J. Yan, Y. Zhang, and K. Fu, "SCRDet: Towards more robust detection for small, cluttered and rotated objects," in *Proc. Int. Conf. Comput. Vis.*, 2019, pp. 8232–8241.

[47] M. Toneva et al., "An empirical study of example forgetting during deep neural network learning," in *Proc. Int. Conf. Learn. Representation*, 2019, pp. 1–18.
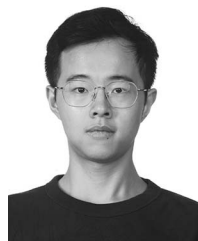
Yue Zhou received the B.S. degree in electronic and information engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2017. He is currently working toward the Eng.D. degree in electronics and information with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China.

He is the leading contributor to the MMRotate open-source project for oriented object detection. His research interests include deep learning, image processing, and radar object detection and recognition.

Xue Jiang (Senior Member, IEEE) received the M.S. degree in acoustics and signal processing from the Chinese Academy of Sciences, Beijing, China, and the Ph.D. degree in electrical engineering from McMaster University, Hamilton, ON, Canada, in 2007 and 2014, respectively.

From 2007 to 2010, she was a Research Engineer with Sunplus Inc., Beijing. From 2014 to 2015, she held a Postdoctoral Fellow position with McMaster University. From 2015 to 2016, she was a Postdoctoral Fellow with Arizona State University, Tempe, AZ, USA. Since 2016, she has been a Professor with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. Her research interests include array processing, robust beamforming, sparse and low-rank learning, data science, and numerical optimization, with applications to radar, sonar, and remote sensing.

Dr. Jiang was one of the finalists of the Best Student Paper Award of the IEEE 2013 Radar Conference. She was the recipient of the McMaster University International Excellence Award in 2012 and 2013. She is also an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING. She serves as the Editorial Board Member for *IET Radar, Sonar and Navigation*.

Zeming Chen received the B.S. degree in automation from the Dalian University of Technology, Dalian, China, in 2020. He is currently working toward the Eng.M. degree in computer technology with Shenzhen International Graduate School, Tsinghua University, Shenzhen, China.

His research interests include semisupervised learning and object detection.

Lin Chen received the B.S. degree in electronic engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, and the Ph.D. degree in information and communication engineering from the Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2017 and 2022, respectively.

He is currently a Senior Engineer with the Wireless Network Research Department, Huawei Technologies Company, Ltd., Shanghai, China. His research interests include machine learning and tensor decomposition.

Xingzhao Liu (Member, IEEE) received the B.S. and M.S. degrees from the Harbin Institute of Technology, Harbin, China, in 1984 and 1992, respectively, and the Ph.D. degree from the University of Tokushima, Tokushima, Japan, in 1995, all in electrical engineering.

From 1984 to 1998, he was an Assistant Professor, an Associate Professor, and a Professor with the Harbin Institute of Technology. Since 1998, he has been a Professor with Shanghai Jiao Tong University, Shanghai, China. His research interests include high-frequency radar and synthetic aperture radar signal processing.