

RSOD: Reliability-Guided Sonar Image Object Detection with Extremely Limited Labels

Chengzhou Li¹, Ping Guo¹, Guanchen Meng¹, Qi Jia^{1*}, Jinyuan Liu¹, Zhu Liu¹, Xiaokang Liu¹, Yu Liu¹, Zhongxuan Luo¹, Xin Fan^{1*}

¹School of Software Technology, Dalian University of Technology
 {lichengzhou, guoping, guanchenmeng, liuzhu, xkliu}@mail.dlut.edu.cn,
 {jiaqi, liuyu882, zxlou, xin.fan}@dlut.edu.cn, atlantis918@hotmail.com

Abstract

Object detection in sonar images is a key technology in underwater detection systems. Compared to natural images, sonar images contain fewer texture details and are more susceptible to noise, making it difficult for non-experts to distinguish subtle differences between classes. This leads to their inability to provide precise annotation data for sonar images. Therefore, designing effective object detection methods for sonar images with extremely limited labels is particularly important. To address this, we propose a teacher-student framework called RSOD, which aims to fully learn the characteristics of sonar images and develop a pseudo-label strategy suitable for these images to mitigate the impact of limited labels. First, RSOD calculates a reliability score by assessing the consistency of the teacher’s predictions across different views. To leverage this score, we introduce an object mixed pseudo-label method to tackle the shortage of labeled data in sonar images. Finally, we optimize the performance of the student by implementing a reliability-guided adaptive constraint. By taking full advantage of unlabeled data, the student can perform well even in situations with extremely limited labels. Notably, on the UATD dataset, our method, using only 5% of labeled data, achieves results that can compete against those of our baseline algorithm trained on 100% labeled data. We also collected a new dataset to provide more valuable data for research in the field of sonar.

Code — <https://github.com/LICZ9/RSOD>

Introduction

Seawater absorbs visible light very strongly, with effective attenuation depths for different wavelengths of only a few meters to a few tens of meters (Xie, Yang, and Qiu 2022). In turbid waters, the deep sea, or night-time environments, visible-light imaging is almost entirely ineffective (Liu et al. 2025; Li et al. 2025b). In contrast, sonar is inherently robust to changes in illumination, suffers a much lower propagation loss, and can detect objects at ranges of tens to hundreds of meters (Cao, Ren, and Sun 2022). As a result, it has been widely adopted for a variety of underwater tasks, including but not limited to accident rescue, equipment maintenance, and biological surveys (Cao, Ren, and Sun 2022; McCann

*Corresponding author.

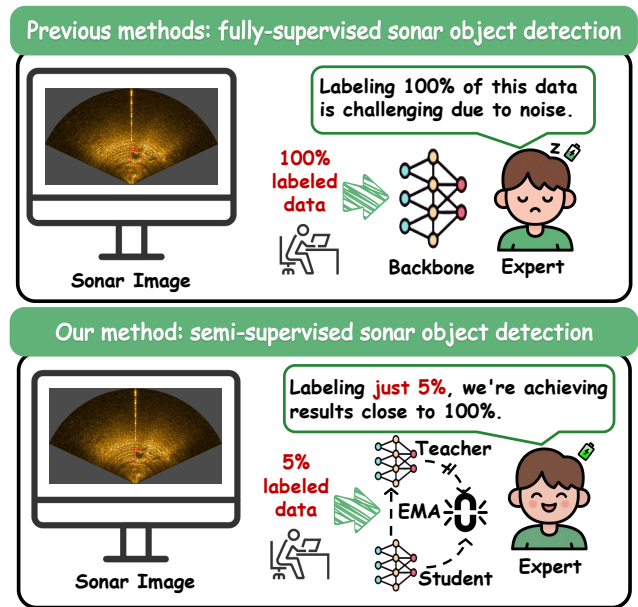


Figure 1: Compared to previous approaches for training sonar image object detection detectors, our method reduces the requirement for labeled data.

et al. 2018). However, despite these advantages, sonar images remain prone to interference from electronic noise, suspended particles, and sediment, resulting in sparse and difficult to interpret object textures (Singh and Valdenegro-Toro 2021; Li 2012). At the same time, the high confidentiality of the sonar data and the steep cost of annotation make it difficult for supervised learning to obtain sufficient labeled data (Li et al. 2025a). Therefore, exploring the intrinsic features of sonar image and developing efficient strategies to mitigate label scarcity are crucial to improving underwater robot object detection performance.

As shown in Fig. 1, to address the issues above, we introduce semi-supervised learning into sonar object detection to leverage unlabeled data and reduce our dependence on manual annotations. Typical semi-supervised methods (Xu et al. 2021; Liu et al. 2024) generate pseudo-labels for unlabeled samples through model predictions, which are then used to

train a student (Hua et al. 2023). Specifically, the teacher first applies weak augmentations (e.g. mild noise, translation, or rotation) to unlabeled sonar images, thereby generating a set of pseudo-labels (Zhang, Pan, and Wang 2022). The student then enforces prediction consistency by using these pseudo-labels to supervise strongly augmented versions of the same images (e.g. random cropping, heavy noise injection).

However, due to the typically limited texture detail information of objects in sonar images, the noise in pseudo-labels increases significantly in scenarios with extremely limited labels, severely restricting the model’s stability and robustness in sonar object detection tasks. Furthermore, while pseudo-labels can significantly enhance the recognition of large objects and dominant classes, they provide very limited assistance for small objects and multi-object scenarios. Therefore, seeking effective methods to reduce pseudo-label noise and designing pseudo-label strategies that balance the detection capabilities of different scales of targets in sonar images are crucial to further improving detection performance in situations with extremely limited labels.

In this paper, we propose RSOD, a semi-supervised learning framework for sonar image object detection aimed at addressing the issue of limited labels. First, we evaluate the predictions of the teacher in multiple augmented views to calculate the reliability score for each pseudo-label bounding box. Next, we select pseudo-labels with high reliability scores and apply our object mixed pseudo-label strategy to effectively reduce the negative impact of pseudo-label noise and balance the model’s learning capability for objects of different scales. Finally, we optimize the student through reliability-guided adaptive constraint, maximizing the utilization of unlabeled data while improving the influence of high quality pseudo-labels on the model.

Our main contributions can be summarized as follows:

- To the best of our knowledge, we are the first to introduce a semi-supervised object detection algorithm specifically designed for sonar images. Our proposed RSOD framework efficiently utilizes unlabeled samples to lower annotation costs while enhancing detector performance.
- We propose a novel method for calculating pseudo-label reliability scores along with an object mixed pseudo-label strategy to maximize the utilization of high quality pseudo-labels and minimize the impact of noisy pseudo-labels on the overall loss.
- We created a new dataset named the Forward-Looking Sonar Image Object Detection (FSOD) dataset, which comprises 3,929 sonar images captured with an Oculus M750d sonar, annotated in 10 distinct object categories.
- Extensive experiments conducted under different settings of labeled data ratios in the UATD and FSOD datasets demonstrate that RSOD outperforms existing state-of-the-art methods, particularly showing significant improvements when only 5% of the data is annotated.

Related Work

Sonar Image Object Detection

In recent years, the rapid advancement of deep learning (Yu et al. 2021; Redmon and Farhadi 2018; Liu et al. 2016)

has dramatically improved the performance of CNN-based sonar image object detection (Tarvainen and Valpola 2017; Valdenegro-Toro 2016), especially Fast R-CNN (Girshick 2015) and Faster R-CNN (Ren et al. 2015). Faster R-CNN introduces a Region Proposal Network (RPN) that directly generates candidate boxes and scores their object in an end-to-end fashion from the raw input, sharing layers with both the classifier and the regressor, and, in doing so, substantially improves detection speed and accuracy. (Wang et al. 2022a) have designed multi-scale convolution structures to enhance the network’s ability to perceive small target features, thereby improving the detection performance of small targets in sonar images. (Shen et al. 2024) incorporate the characteristics of underwater sonar images, such as grayscale, low contrast, and noise interference, to optimize the performance of sonar object detectors.

However, these methods do not take into account the label scarcity problem that sonar images typically face in real-world scenarios. We introduce a teacher-student framework to fully leverage unlabeled data, thus mitigating the impact of insufficient labeled data on the model.

Semi-Supervised Object Detection

Sonar images, because of their high confidentiality and the inherent difficulty of interpreting overlapping echoes, suffer from an extreme scarcity of labeled data. Semi-supervised object detection (SSOD) seeks to address this by cleverly integrating potential object information from unlabeled images to boost detection performance (Wang et al. 2025; Liu et al. 2023), typically through a teacher-student paradigm: a teacher is first trained on the limited labeled set (Liu et al. 2021; Wang et al. 2023); It then generates pseudo-labels for the unlabeled images, which are used to jointly train the student model. To improve pseudo-label quality, researchers have incorporated consistency regularization (Liu, Ma, and Kira 2022), confidence threshold filtering (Xu et al. 2021), and random data increase perturbations (Chen et al. 2023).

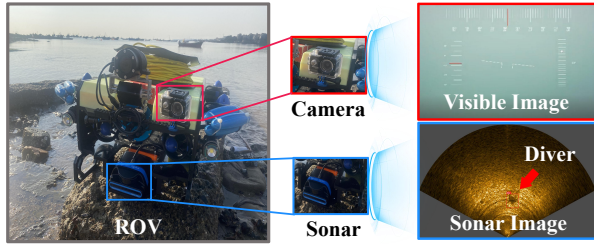
Existing SSOD methods cannot effectively address the issue of noisy pseudo-labels generated by the teacher due to the limited texture detail in sonar images. Therefore, we propose a new reliability-guided semi-supervised object detection method for sonar images that reevaluates all pseudo-labels using reliability scores, thereby effectively mitigating the impact of noisy pseudo-labels on the detector.

Data Construction

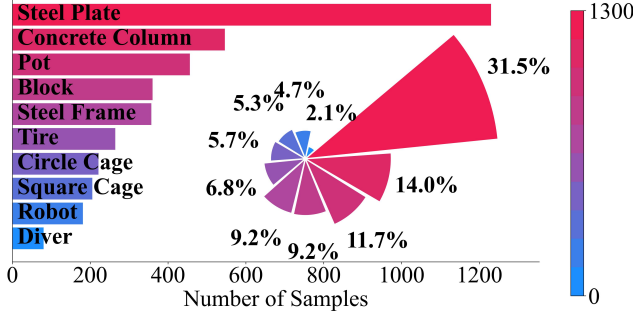
Underwater Sonar Data Collection

The scarcity of publicly available sonar image datasets significantly hampers progress in sonar object detection. This lack of resources limits the development of advanced algorithms and restricts research on the features of sonar images and the generalization of algorithms.

To address the aforementioned challenges, we present a new dataset, the Forward-Looking Sonar Image Object Detection (FSOD) dataset. Data were collected in Bohai Bay, as shown in Fig. 2(a). Specifically, we used a remotely operated vehicle (ROV) equipped with a multibeam forward-looking sonar (Oculus M750d) to acquire underwater data.



(a) Illustration of sonar data collection platform.



(b) Sample distribution in the FSOD dataset.

Figure 2: (a) The collection process of the FSOD dataset. (b) The number in each category of the FSOD dataset.

To reduce mutual interference between objects and facilitate location, the objects were connected and suspended by ropes at depths of approximately 3 to 20 meters below the water surface and placed at fixed intervals. During the data collection process, each object was imaged from different angles and distances (2–15 meters), thus enriching the diversity of the dataset. The collected objects were then filtered, pre-processed, and annotated, resulting in a dataset comprising 3,929 images across ten categories. Fig. 2(b) shows the number of each category in the dataset. In particular, the steel frame category contains the most samples with 1,231 images, accounting for 31.5% of the dataset, while the diver category has the fewest with only 81 samples, accounting for 2.1%, demonstrating the inherent long-tailed distribution of the FSOD dataset. It should be mentioned that, compared to the mannequin samples provided in the UATD dataset (Xie, Yang, and Qiu 2022), the FSOD dataset also includes a diver category, offering important data resources for the development of underwater emergency rescue, accident prevention, and diving safety assurance systems.

Methodology

Problem Definition

The goal of RSOD algorithm is to learn a detector to locate and identify objects. The training dataset consists of two parts: a set of labeled sonar images $D_l = \{S_i^l, Y_i^l\}_{i=1}^{N_l}$ and a set of unlabeled sonar images $D_u = \{S_i^u\}_{i=1}^{N_u}$. Here, N_l and N_u denote the numbers of labeled and unlabeled sonar images, respectively. For each labeled sonar image S^l , the corresponding annotation Y^l provides the locations, sizes, and categories of all the boxes.

Overview

As shown in Fig. 3, the overall framework of RSOD. (a) Illustrating the basic structure of the method, (b) the process of assessing the reliability of pseudo-labels, (c) the generation process of object mixed pseudo-labels, and (d) the loss optimization process by reliability-guided adaptive constraint. In Fig. 3(a), the proposed RSOD framework consists of two detection networks with identical architectures, namely the teacher network and the student network. The teacher network is responsible for generating pseudo-labels, while the student network is jointly trained using both these pseudo-labels and the ground-truth labels. Based on this process, the goal of optimizing the student network can be formulated as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \lambda_u \cdot \mathcal{L}_{unsup}, \quad (1)$$

where λ_u is the hyper-parameter to adjust the contribution of the unsupervised loss.

For labeled data $D_l = \{S_i^l, Y_i^l\}_{i=1}^{N_l}$, the supervised loss \mathcal{L}_{sup} for object detection consists of the classification loss \mathcal{L}_{cls} , the regression loss \mathcal{L}_{reg} (Ren et al. 2015).

$$\mathcal{L}_{sup} = \sum_i \mathcal{L}_{cls}(S_i^l, Y_i^l) + \mathcal{L}_{reg}(S_i^l, Y_i^l). \quad (2)$$

The unsupervised loss \mathcal{L}_{unsup} will be discussed in the chapter on Reliability-Guided Adaptive Constraint.

Pseudo-Label Reliability Assessment (PLRA)

In semi-supervised object detection, the performance enhancements of the student hinge on high quality pseudo-labels generated by the teacher. Current approaches often filter pseudo-labels based solely on a classification confidence threshold, neglecting the precision of bounding-box localization. Consequently, they end up introducing numerous boxes that are high confidence but coarse or even high confidence but incorrect as pseudo-labels.

As shown in Fig. 3(b), we propose a pseudo-label reliability assessment scheme based on multi-view consistency. Specifically, for an unlabeled sonar image S^u , we first perform multi-view augmentation operations, including mirrored and scaling operations. The teacher network will predict the bounding boxes for S^u and its augmented view after augmentation, denoted as Y^{uo} and Y^{ua} , respectively. Next, we match each bounding box in Y^{uo} with the corresponding bounding box in Y^{ua} according to their IoU values. The boxes in the bounding of the various augmented views that match those of the original view are denoted as (y^{uo}, y^{ua}) .

First, for a given $\{y_j^{uo}\}_{j=1}^{N_j}$, we search for $\{y_p^{ua}\}_{p=1}^{N_p}$ in Y^{ua} that has the highest IoU denoted as IoU^* with y_j^{uo} . The computation is formulated as follows:

$$IoU^* = \max_{y_j^{uo} \in Y^{uo}, y_p^{ua} \in Y^{ua}} IoU(y_j^{uo}, y_p^{ua}), \quad (3)$$

where N_j and N_p represent the number of pseudo-label boxes in the original view and the augmented view, respectively. After identifying the box y_p^{ua} that most closely matches the original view box y_j^{uo} in the augmented views,

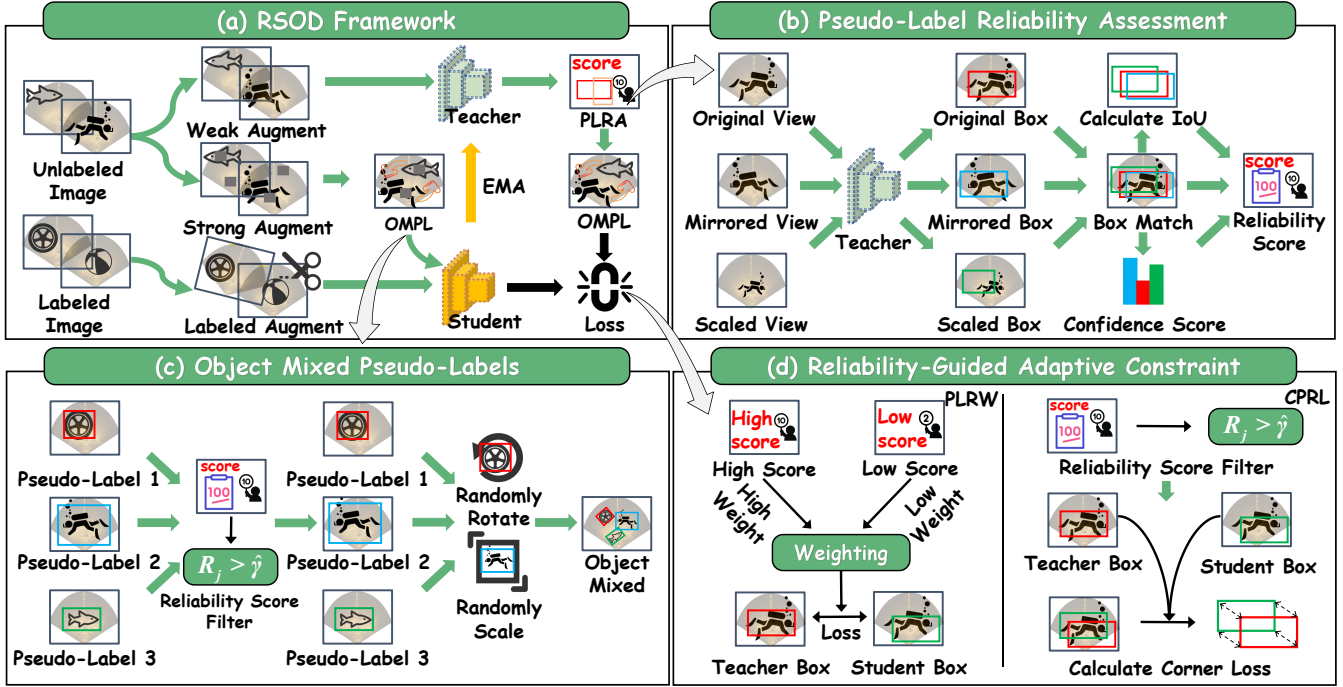


Figure 3: The overall architecture of RSOD. (a) Illustrating the basic structure of the RSOD, (b) the process of assessing the reliability of pseudo-labels, (c) the generation process of object mixed pseudo-labels, and (d) the loss optimization process by reliability-guided adaptive constraint.

we comprehensively assess the reliability of the pseudo-labels generated by the teacher. To achieve this, we propose a reliability score R_j that jointly considers both the positional consistency and the category consistency of the predicted boxes across various augmented views. The calculation of the reliability score is as follows:

$$R_j = \text{Sigmoid} \left(\frac{1}{H} \cdot \frac{\sum_{a \in \mathcal{A}} [IoU^* \cdot (1 - |s_j^o - s_p^a|) \cdot \beta]}{N_a} \right), \quad (4)$$

where H is the reliability scaling factor, \mathcal{A} denotes the set of augmented views, including mirrored and scaling operations. s_j^o represents the class confidence score of the predicted box y_j^{uo} from the original view, while s_p^a denotes the class confidence of the predicted box y_p^{ua} . N_a is the number of augmented views, which is set to 2, including mirrored and scaled views. The main purpose of β is to determine whether the predicted class of the sonar image S^u pseudo-label bounding box y_j^{uo} is consistent with that of the corresponding augmented view pseudo-label bounding box y_p^{ua} . Its calculation formula is as follows:

$$\beta = \begin{cases} 1, & \text{if } c_j^o = c_p^a, \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

where c_j^o denotes the class of the bounding box y_j^{uo} and c_p^a denotes the class of the bounding box y_p^{ua} .

By comparing the teacher's predictions across different augmented views, we calculate a reliability score for each

pseudo-label bounding box. Pseudo-labels that demonstrate higher consistency in predictions across different views will receive higher scores.

Object Mixed Pseudo-Labels (OMPL)

To fully utilize the high reliability pseudo-labels provided by the teacher and enhance the model's ability to recognize challenging samples (such as multiple small objects), we propose an object mixed pseudo-label strategy. This strategy involves rotating and scaling the pseudo-label bounding boxes with high reliability scores and mixing them with pseudo-labels from other samples to generate more pseudo-labels containing small and multiple objects. These pseudo-label bounding boxes can cover a substantial amount of irrelevant noise in the background of sonar images, thereby maximizing the utilization of valuable high reliability object pseudo-label bounding boxes.

As shown in Fig. 3(c), we propose the object mixed pseudo-label strategy to fully leverage the high confidence pseudo-labels provided by the teacher. First, we obtain the set of pseudo-label bounding boxes $\hat{y}_{S_k}^u = \{y_{S_k}^u \mid R_j > \hat{\gamma}, k \in \{2, 3, \dots, K\}\}$ with high reliability scores from the k -th image in the current batch. $\hat{\gamma}$ is the reliability score threshold, and $y_{S_k}^u$ denotes the set of pseudo-label bounding boxes for the k -th image in the current batch. R_j represents the reliability score of the pseudo-label bounding boxes $y_{S_k}^u$. We can paste the augmented pseudo-label predicted boxes back onto the first image of the current

batch. The operation can be described as follows:

$$S'_1 = S_1 + \sum_{b_{S_k} \in \hat{y}_{S_k}^u} \bar{F}_b \cdot P(S_1, b_{S_k}), \quad (6)$$

where S_1 represents the first sonar image in the current batch and S'_1 represents the updated image after the paste operation is completed. $\hat{y}_{S_k}^u$ represents $\hat{y}_{S_k}^u$ after augmentation (the augmentation operations include random rotation and scaling of the pseudo-labels). $P(S_1, b_{S_k})$ represents pasting the pseudo-label boxes b_{S_k} into the first image S_1 . b_{S_k} represents a bounding box in $\hat{y}_{S_k}^u$. \bar{F}_b is the object selection flag, which can be expressed as:

$$\bar{F}_b = \begin{cases} 1, & \text{if IoU}(b_{S_1}, b_{S_k}) < \delta, \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

Only when the IoU between b_{S_k} and any pseudo-label box b_{S_1} in the first image S_1 is less than the IoU threshold δ that we set will the paste operation be performed to avoid the pasted pseudo-label boxes overlapping.

Reliability-Guided Adaptive Constraint

As shown in Fig. 3(d), for the unlabeled data $D_u = \{S_i^u\}_{i=1}^{N_u}$, we utilize a reliability-guided approach to compute the unsupervised loss, adjusting the influence of each pseudo-label on the overall loss.

Pseudo-Label Reliability Weighting (PLRW) We use reliability scores to differentiate the extent to which different pseudo-labels impact the loss function. We first evaluate the pseudo-label boxes provided by the teacher using the PLRA, obtaining a score for each pseudo-label box. The computed scores are then used to weight the pseudo-label boxes, so that boxes with higher reliability scores from the teacher are assigned greater weights when calculating the loss with the student. The loss \mathcal{L}_{PLRW} guided by the reliability score is calculated as follows:

$$\mathcal{L}_{PLRW} = \frac{1}{N_j} \sum_{j=1}^{N_j} R_j \cdot (l_{cls}(b_j^u, T^u) + l_{reg}(b_j^u, T^u)), \quad (8)$$

where T^u denotes the set of pseudo-boxes generated by the teacher, l_{cls} is the loss in box classification and l_{reg} is the loss in box regression (Xu et al. 2021). N_j represents the number of candidate boxes predicted by the student. b_j^u denotes the j -th predicted box in the set of candidate boxes predicted by the student, and R_j denotes the reliability score of the pseudo-box of the teacher corresponding to the bounding box of the student.

Corner Point Regression Loss (CPRL) To further leverage pseudo-labels with high reliability scores from the teacher, we enhance the precision of the student’s bounding box localization by introducing a corner point regression loss. For each predicted bounding box $y_j^u \in T^u$ from the teacher, R_j is the reliability score corresponding to y_j^u . If $R_j > \hat{\gamma}$, we apply corner point regression loss to strengthen the supervision of the teacher over the box location predictions of the student. The loss of corner point regression loss

\mathcal{L}_{CPRL} is as follows:

$$\mathcal{L}_{CPRL} = \frac{1}{N_j} \sum_j^{N_j} \frac{\|C_j^T - C_j^S\|_F^2}{l_j^2}, \quad (9)$$

where N_j denotes the number of boxes that both exceed the teacher’s reliability score threshold $\hat{\gamma}$ and match the predicted boxes of the student. C_j^T denotes all corner points corresponding to the predicted boxes from the teacher, which can be expressed as a two-dimensional matrix, where each row denotes the 2D coordinates of a corner point (Putra, Murni, and Chahyati 2025). C_j^S represents the corner coordinates of the predicted box from the student. $\|\cdot\|_F$ represents the Frobenius norm, which is the square root of the sum of the squares of all the elements in a matrix. l_j^2 represents the diagonal length of the minimum enclosing rectangle of the predicted matched boxes from the teacher and student.

The final unsupervised loss \mathcal{L}_{unsup} can be expressed as:

$$\mathcal{L}_{unsup} = \mathcal{L}_{PLRW} + \mathcal{L}_{CPRL}. \quad (10)$$

Experiment

Datasets and Experiment Settings

Dataset This study conducted comparison and ablation experiments on the FSOD dataset we collected and the publicly available UATD dataset (Xie, Yang, and Qiu 2022). The UATD dataset is widely used for object detection tasks in underwater sonar images (Yang et al. 2024; Wang et al. 2022b). We utilized the original partition of the dataset, with 7,600 images in the training set and 800 images each in the test and validation sets. Since there is no current research on semi-supervised object detection in sonar images, we follow the protocols used for semi-supervised object detection (Wang et al. 2023). In practice, we randomly selected 1%, 5%, and 10% of the images in the training set as labeled data, while the remaining images were treated as unlabeled data.

For the FSOD dataset that we collected, we divided it into training, validation, and test sets in a 6:2:2 ratio. This dataset exhibits a long-tail distribution, with only 81 images belonging to the diver category. We randomly selected 5%, 10%, and 20% of the images as labeled data, and the rest were considered unlabeled data.

Evaluation Metrics For all experiments, we evaluated the performance on the UATD and FSOD datasets, reporting the results using standard mean average precision (mAP) as an evaluation metric (Xu et al. 2021; Chen et al. 2023).

Implementation Details To ensure a fair comparison, we follow STAC (Sohn et al. 2020) to use FasterRCNN with FPN (Lin et al. 2017) and ResNet-50 backbone (He et al. 2016) as our object detector, where the feature weights are initialized by the ImageNet-pretrained model, same as existing works (Liu et al. 2021; Jeong et al. 2019). Furthermore, an asymmetric data augmentation strategy is applied to the unlabeled data, including weak augmentation by random flipping (Li et al. 2022) and strong augmentation through a greater degree of random flipping and random Gaussian blur (Tang et al. 2021). The entire model is trained for 180K

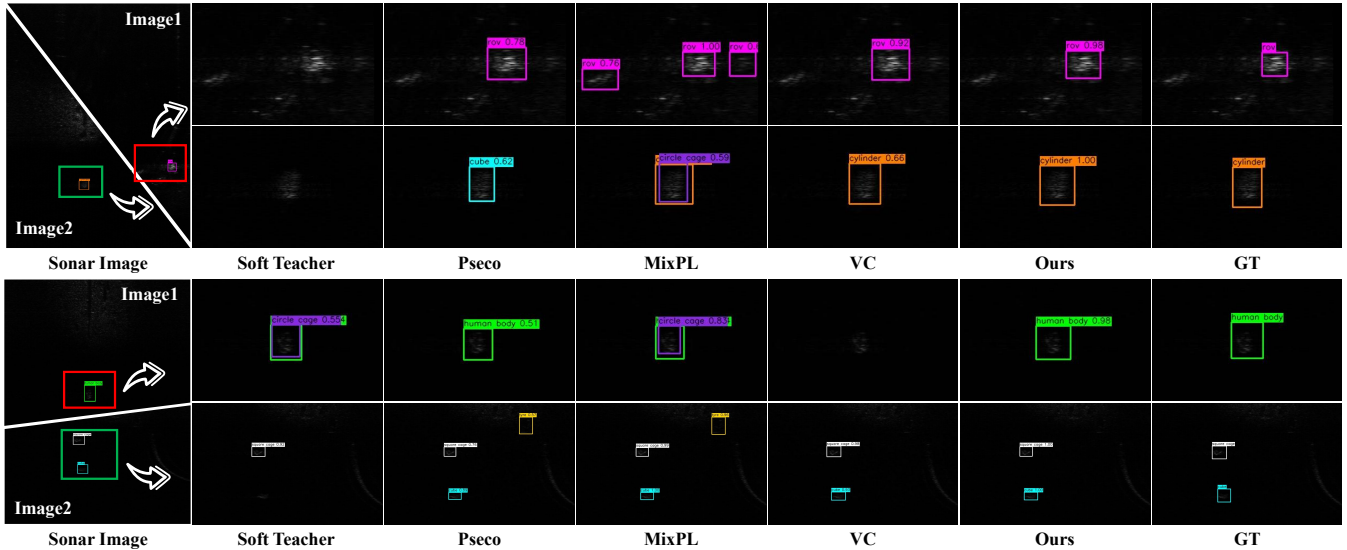


Figure 4: Qualitative comparison of state-of-the-art object detection methods on the UATD dataset with 5% labeled data.

Method	Dataset	UATD			FSOD		
		1%	5%	10%	5%	10%	20%
Faster R-CNN (Ren et al. 2015)(Supervised)		13.7	22.9	26.2	44.9	50.6	59.6
Soft Teacher (Xu et al. 2021)		<i>23.0(+9.3)</i>	30.8(+7.9)	33.1(+6.9)	56.4(+11.5)	65.7(+15.1)	72.2(+12.6)
Unbiased Teacher V2 (Liu, Ma, and Kira 2022)		<i>21.6(+7.9)</i>	31.0(+8.1)	33.4(+7.2)	65.0(+20.1)	70.1(+19.5)	72.8(+13.2)
Pseco (Li et al. 2022)		22.9(+9.2)	<i>32.7(+9.8)</i>	<i>33.6(+7.4)</i>	65.7(+20.8)	<i>70.5(+19.9)</i>	<i>74.6(+15.0)</i>
MixPL (Chen et al. 2023)		22.2(+8.5)	<i>31.6(+8.7)</i>	<i>32.2(+6.0)</i>	64.3(+19.4)	<i>70.2(+19.6)</i>	<i>74.5(+14.9)</i>
VC (Chen, Han, and Debattista 2024)		21.5(+7.8)	30.9(+8.0)	32.2(+6.0)	<i>65.9(+21.0)</i>	69.6(+19.0)	74.0(+14.4)
RSOD(Ours)		23.2(+9.5)	34.7(+11.8)	34.8(+8.6)	67.4(+22.5)	72.6(+22.0)	75.5(+15.9)

Table 1: Quantitative comparisons on the UATD dataset and FSOD dataset under the mAP. Experiments on the UADT dataset were conducted under labeled data settings of 1%, 5%, and 10%, while experiments on the FSOD dataset were conducted under labeled data settings of 5%, 10%, and 20%. The best results are shown in **bold**, and the second best results are shown in *italic*.

iterations on a single NVIDIA RTX 4090 GPU. We use SGD optimizer for all detectors with a learning rate of 0.001, a momentum of 0.9, and a weight decay of 0.0001.

Main Results

In this section, we compare our RSOD with two-stage classical fully supervised object detectors, Faster R-CNN (Ren et al. 2015). Furthermore, we compare it with various state-of-the-art semi-supervised object detection (SSOD) methods, including Unbiased Teacher V2 (Liu, Ma, and Kira 2022), Soft Teacher (Xu et al. 2021), MixPL (Chen et al. 2023), Pseco (Li et al. 2022), and VC (Chen, Han, and Debattista 2024). Using publicly available source code, we have reimplemented these benchmarks on the proposed UATD and FSOD datasets.

Qualitative Comparison As shown in Fig. 4, in the detection of small objects in underwater sonar images, both the Soft Teacher and VC methods experience missed detections, while our approach successfully identifies the objects with high confidence. This indicates that our object

Method	UATD		
	1%	5%	10%
Faster R-CNN	2.9	7.1	7.3
Soft Teacher	<i>9.9(+7.0)</i>	11.2(+4.1)	12.2(+4.9)
Unbiased Teacher V2	<i>9.1(+6.2)</i>	10.4(+3.3)	<i>12.6(+5.3)</i>
Pseco	8.9(+6.0)	12.0(+4.9)	12.3(+5.0)
MixPL	9.1(+6.2)	10.1(+3.0)	11.4(+4.1)
VC	5.7(+2.8)	<i>12.2(+5.1)</i>	12.4(+5.1)
RSOD (Ours)	10.1(+7.2)	12.6(+5.5)	12.8(+5.5)

Table 2: Quantitative comparisons on the UATD dataset under the mAP(small) with different labeled data settings.

mixed pseudo-label algorithm significantly enhances the detection ability for small objects. In contrast, both the Pseco and MixPL methods produce false detections, and their predicted bounding boxes have high confidence levels. Our method remains consistent with ground truth, highlighting that our reliability scoring assessment strategy has been ef-

Model	EMA	PLRW	CPRL	OMPL	mAP
M1	-	-	-	-	44.9
M2	✓	-	-	-	64.3
M3	✓	✓	-	-	65.2
M4	✓	✓	✓	-	66.3
M5	✓	✓	✓	✓	67.4

Table 3: Ablation study on key components of RSOD conducted on the FSOD dataset with 5% labeled data.

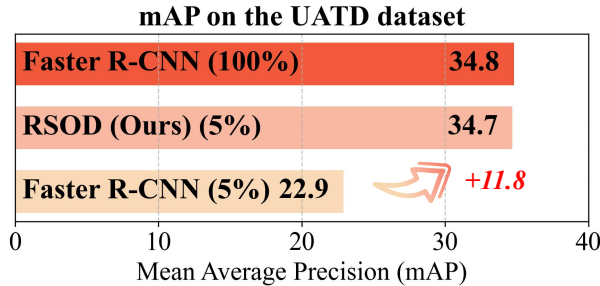


Figure 5: mAP on the UATD dataset with 5% labeled data.

fective in suppressing the influence of unstable pseudo-label boxes. Our method demonstrates overall performance that exceeds other approaches. Not only does it avoid missed detections and false detections, but it also provides high confidence for all correctly detected objects.

Quantitative Comparison As shown in Table 1, RSOD significantly outperforms other methods in the UATD and FSOD datasets, demonstrating the effectiveness of our proposed reliability score evaluation method and the object mixed pseudo-label algorithm. This approach notably increases the utilization of high reliability pseudo-labels while effectively suppressing the impact of unreliable pseudo-labels, successfully enhancing the predictive accuracy of the student. As shown in Table 2, RSOD also improves the ability of the model to predict small objects.

As shown in Fig. 5, on the UATD dataset, when the proportion of labeled data is only 5%, the mAP of RSOD can compete against the fully supervised results obtained from the Faster R-CNN baseline detector using training data 100%. This effectively alleviates the negative impact of scarce annotations on model performance.

Ablation Study

In this section, we conducted extensive research to validate our key design and critical parameter settings on the FSOD dataset with 5% labeled data.

Component Analysis The results in Table 3 validate the effectiveness of key components in RSOD. M1 performed better with the Faster R-CNN detector that we used under fully supervised conditions. M2 indicates the results after introducing the teacher-student network. M3 shows the performance of the model after incorporating PLRW. M4 illus-

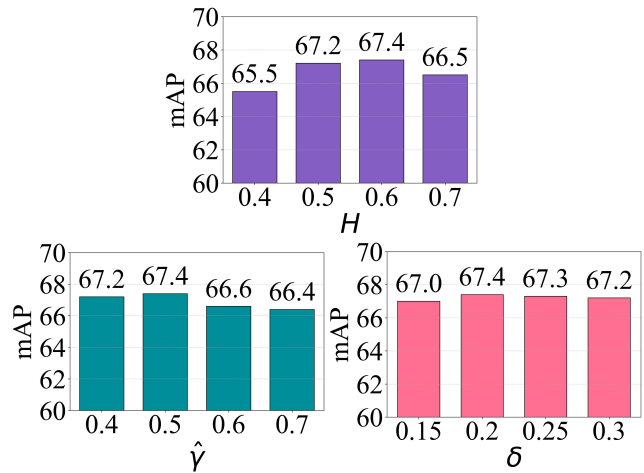


Figure 6: mAP on the FSOD dataset with 5% labeled data.

trates the model’s performance following the incorporation of CPRL. M5 refers to the performance after applying the OMPL. From the experimental results, it can be seen that the introduction of the designed teacher-student network architecture has played a significant role. Guided by the reliability scores of the pseudo-label algorithm, both PLRW and CPRL effectively reduce the noise in the pseudo-label bounding boxes, greatly improving the mAP of the student. Our proposed object mixed pseudo-label strategy effectively enhances the model’s detection capabilities for challenging samples, such as small objects and multiple objects.

Parameter Analysis As shown in Fig. 6, we analyzed the key parameters during the experimental process, including the reliability scaling factor H , the reliability score threshold $\hat{\gamma}$, and the IoU threshold δ . The experimental results indicate that setting H to 0.6 effectively expands the gap in reliability scores between pseudo-labels, helping the model better distinguish between high-quality and low-quality pseudo-labels. Furthermore, when the reliability score threshold $\hat{\gamma}$ is set to 0.5 and the IoU threshold δ is set to 0.2, the model achieves the best result of 67.4. This indicates that the object mixed pseudo-label strategy and corner point regression loss assist the student detector in achieving optimal performance.

Conclusion

In this paper, we combine the teacher-student architecture with pseudo-label reliability assessment methods to effectively utilize unlabeled data. The student is better able to distinguish and learn different pseudo-labels with the support of reliability scores, significantly mitigating the negative impact of noisy pseudo-labels. Additionally, the object mixed pseudo-label strategy effectively enhances the model’s ability to handle difficult samples, such as small and multiple targets. We conducted extensive experiments on two datasets, and the results demonstrate that RSOD exhibits significant superiority over other methods in various scenarios with limited labels.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant Nos. 62495085, 62027826, 62442603, 62272083, 62472066 and 62495083 and by the Key Research and Development Program of Liaoning Province, under No. 2023JH26/10200014.

References

- Cao, X.; Ren, L.; and Sun, C. 2022. Dynamic target tracking control of autonomous underwater vehicle based on trajectory prediction. *IEEE Transactions on Cybernetics*, 53(3): 1968–1981.
- Chen, C.; Han, J.; and Debattista, K. 2024. Virtual category learning: A semi-supervised learning method for dense prediction with extremely limited labels. *IEEE transactions on pattern analysis and machine intelligence*, 46(8): 5595–5611.
- Chen, Z.; Zhang, W.; Wang, X.; Chen, K.; and Wang, Z. 2023. Mixed pseudo labels for semi-supervised object detection. *arXiv preprint arXiv:2312.07006*.
- Girshick, R. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 1440–1448.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Hua, W.; Liang, D.; Li, J.; Liu, X.; Zou, Z.; Ye, X.; and Bai, X. 2023. Sood: Towards semi-supervised oriented object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 15558–15567.
- Jeong, J.; Lee, S.; Kim, J.; and Kwak, N. 2019. Consistency-based semi-supervised learning for object detection. *Advances in neural information processing systems*, 32.
- Li, C.; Liu, X.; Jia, Q.; Liu, J.; Jiang, Z.; Feng, L.; Liu, Y.; Luo, Z.; and Fan, X. 2025a. Physics-Guided Sonar Image Fine-grained Recognition under Scarce Annotations. In *Proceedings of the 33rd ACM International Conference on Multimedia*, 1356–1365.
- Li, G.; Li, X.; Wang, Y.; Wu, Y.; Liang, D.; and Zhang, S. 2022. Pseco: Pseudo labeling and consistency training for semi-supervised object detection. In *European Conference on Computer Vision*, 457–472. Springer.
- Li, Q. 2012. *Digital sonar design in underwater acoustics: principles and applications*. Springer Science & Business Media.
- Li, X.; Wang, Z.; Zou, Y.; Chen, Z.; Ma, J.; Jiang, Z.; Ma, L.; and Liu, J. 2025b. Difisir: A diffusion model with gradient guidance for infrared image super-resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 7534–7544.
- Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; and Szeliski, S. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125.
- Liu, C.; Zhang, W.; Lin, X.; Zhang, W.; Tan, X.; Han, J.; Li, X.; Ding, E.; and Wang, J. 2023. Ambiguity-resistant semi-supervised learning for dense object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 15579–15588.
- Liu, J.; Zhang, B.; Mei, Q.; Li, X.; Zou, Y.; Jiang, Z.; Ma, L.; Liu, R.; and Fan, X. 2025. DCEvo: Discriminative Cross-Dimensional Evolutionary Learning for Infrared and Visible Image Fusion. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2226–2235.
- Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; and Berg, A. C. 2016. Ssd: Single shot multibox detector. In *European conference on computer vision*, 21–37. Springer.
- Liu, Y.; Feng, L.; Jia, Q.; Liu, Z.; and Cao, Z.-H. 2024. Two Teachers Are Better Than One: Semi-supervised Elliptical Object Detection by Dual-Teacher Collaborative Guidance. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 6355–6363.
- Liu, Y.; Ma, C.; He, Z.; Kuo, C.; Chen, K.; Zhang, P.; Wu, B.; Kira, Z.; and Vajda, P. 2021. Unbiased Teacher for Semi-Supervised Object Detection. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*.
- Liu, Y.-C.; Ma, C.-Y.; and Kira, Z. 2022. Unbiased teacher v2: Semi-supervised object detection for anchor-free and anchor-based detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9819–9828.
- McCann, E.; Li, L.; Pangle, K.; Johnson, N.; and Eickholt, J. 2018. An underwater observation dataset for fish classification and fishery assessment. *Scientific data*, 5(1): 1–8.
- Putra, H. A. A.; Murni, A.; and Chahyati, D. 2025. Enhancing Bounding Box Regression for Object Detection: Dimensional Angle Precision IoU-Loss. *IEEE Access*.
- Redmon, J.; and Farhadi, A. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.
- Shen, C.; Shen, L.; Li, M.; and Yu, M. 2024. EPL-UFLSID: Efficient Pseudo Labels-Driven Underwater Forward-Looking Sonar Images Object Detection. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 4349–4357.
- Singh, D.; and Valdenegro-Toro, M. 2021. The marine debris dataset for forward-looking sonar semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 3741–3749.
- Sohn, K.; Zhang, Z.; Li, C.-L.; Zhang, H.; Lee, C.-Y.; and Pfister, T. 2020. A simple semi-supervised learning framework for object detection. *arXiv preprint arXiv:2005.04757*.
- Tang, Y.; Chen, W.; Luo, Y.; and Zhang, Y. 2021. Humble teachers teach better students for semi-supervised object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3132–3141.

Tarvainen, A.; and Valpola, H. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30.

Valdenegro-Toro, M. 2016. Objectness scoring and detection proposals in forward-looking sonar images with convolutional neural networks. In *IAPR workshop on artificial neural networks in pattern recognition*, 209–219. Springer.

Wang, C.; Xu, C.; Li, X.; Li, Y.; Guo, X.; Gu, Z.; and Cui, Z. 2025. Multi-clue consistency learning to bridge gaps between general and oriented object in semi-supervised detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 7582–7590.

Wang, J.; Feng, C.; Wang, L.; Li, G.; and He, B. 2022a. Detection of weak and small targets in forward-looking sonar image using multi-branch shuttle neural network. *IEEE Sensors Journal*, 22(7): 6772–6783.

Wang, X.; Yang, X.; Zhang, S.; Li, Y.; Feng, L.; Fang, S.; Lyu, C.; Chen, K.; and Zhang, W. 2023. Consistent-teacher: Towards reducing inconsistent pseudo-targets in semi-supervised object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3240–3249.

Wang, Z.; Guo, J.; Zeng, L.; Zhang, C.; and Wang, B. 2022b. MLFFNet: Multilevel feature fusion network for object detection in sonar images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–19.

Xie, K.; Yang, J.; and Qiu, K. 2022. A dataset with multi-beam forward-looking sonar for underwater object detection. *Scientific Data*, 9(1): 739.

Xu, M.; Zhang, Z.; Hu, H.; Wang, J.; Wang, L.; Wei, F.; Bai, X.; and Liu, Z. 2021. End-to-end semi-supervised object detection with soft teacher. In *Proceedings of the IEEE/CVF international conference on computer vision*, 3060–3069.

Yang, H.; Zhou, T.; Jiang, H.; Yu, X.; and Xu, S. 2024. A lightweight underwater target detection network for forward-looking sonar images. *IEEE Transactions on Instrumentation and Measurement*.

Yu, Y.; Zhao, J.; Gong, Q.; Huang, C.; Zheng, G.; and Ma, J. 2021. Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOv5. *Remote Sensing*, 13(18): 3555.

Zhang, F.; Pan, T.; and Wang, B. 2022. Semi-supervised object detection with adaptive class-rebalancing self-training. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 3252–3261.