Response Letter

Dear Reviewer,

Re: Manuscript ID: remotesensing-3046409 and Title: "DATA MATTERS: Rethinking Data Distribution in Semi-Supervised Oriented SAR Ship Detection".

Thank you for your encouragement and comments concerning our manuscript. Those comments are valuable and very helpful. We have read through comments carefully and have made corrections as well as responded to each one individually. Revisions in the text are shown using red highlight and the responses to your comments are presented following.

We highly appreciate your time and consideration.

Best regards.

Yours sincerely,
Yimin Yang
Department of Electronic Engineering, Tsinghua University, Beijing 100084,
yangym22@mails.tsinghua.edu.cn

Minor review:

Q1: TITLE: please, check the word "oriented". It is wrongly written.

We sincerely apologize for this serious oversight. We have corrected the spelling error in the title, changing 'Oriented' to the correct form.

Q2: Figure 2: Where is the end of the process? Is it the Equation 10?

Thank you very much for your valuable feedback. We have made the following modifications: As you mentioned, the end of the process shown in Figure 2 is the Equation 10. However, considering that the original figure was not easy to understand, we have redrawn Figure 2 on page 2. Below is the newly drawn schematic diagram: the flow direction is from left to right, unlike the previous chaotic layout, and we have clearly marked the inputs and outputs.

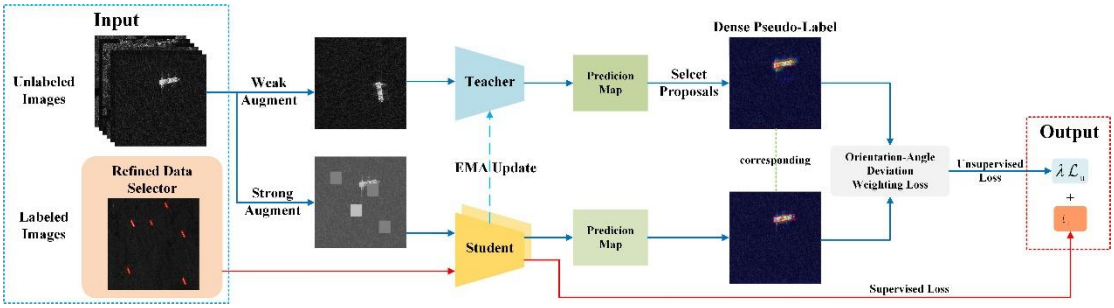


Figure 2. The pipeline of our proposed framework for semi-supervised oriented SAR ship detection.Each input batch contains both labeled and unlabeled data, with labeled data selected offline viarefined data selector (RDS). In the unsupervised training part, the teacher model uses weakly aug-mented data, while the student model uses strongly, where only the student model is trained andthe teacher model is updated through the exponential moving average (EMA). The unsupervisedloss is calculated by combining the prediction maps of the teacher model and the student model,where the bounding box loss is Gaussian Wasserstein distance (GWD) loss, and then weighted bythe orientation-angle deviation. The supervised training loss is calculated based on the differencebetween the ground truth and the student model's predictions on labeled data. The overall loss isobtained by weighting and summing the supervised training loss and the unsupervised training loss.

Q3: Figure 7: Despite it being depicted in a line that saves space, one can see it and analyze it better if it could be depicted in two lines in a bigger size.

Thank you very much for your valuable feedback. We have made the following modifications: We have adjusted the layout of Figure 7, changing it from a single row to two rows. Since an additional figure was added earlier, the modified figure is now Figure 8 on page 12.

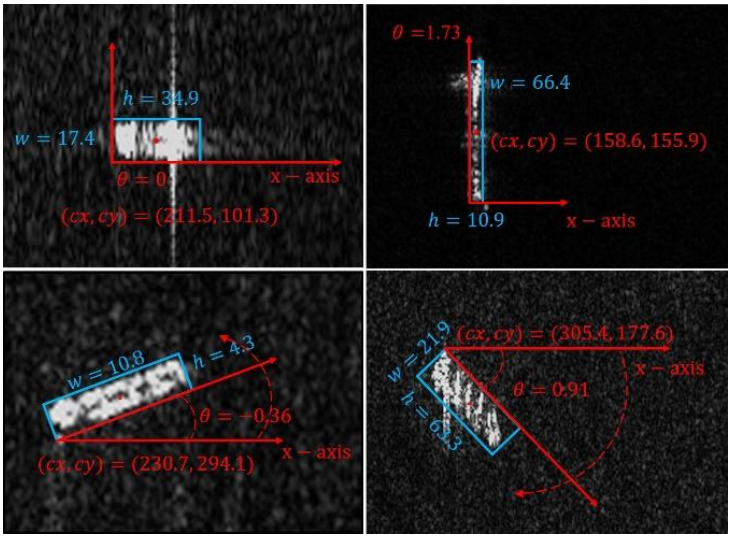


Figure 8. Schematic diagram of the Long Edge definition method used in the RSDD-SAR dataset.

Q4: Figure 8: Authors could enhance the visualization of some challenging cases depicted. Also, you could discuss some aspects of the distribution graph and the correspondent SAR image.

Thank you very much for your valuable feedback. We have made the following modifications: To achieve better visualization, we have added a row in Figure 8 to display the ships in the original SAR images. The modified figure is now Figure 9 on page 15. Additionally, we have included further explanations about the image in lines 439-445 on page 15:

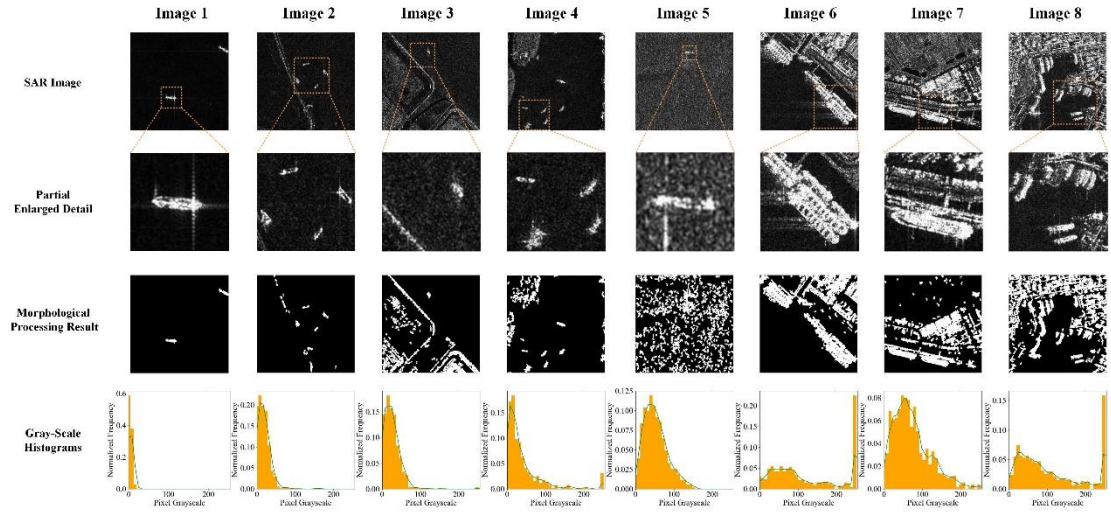


Figure 9. There are eight representative sample images, partial enlarged details, morphologicalresults and their gray-scale histograms.

" To achieve better visualization, as shown in second row in Figure 9, the images are locally enlarged to reveal more details of the ships. These phenomena reflect that as

the images progress from left to right, the mean and variance of the gray-scale generally increase. This is manifested in the histogram as the peak position and peak width rise accordingly, as shown in forth row Figure 9. From the results of morphological processing, as shown in third row Figure 9, we can observe that the spatial distribution of pixels becomes increasingly complex: the connected regions obtained after processing increase in number, size, and density. "

Q5: Figure 11: Same comments as in Figure 8.

Thank you very much for your valuable feedback. We have made the following modifications: To achieve better visualization for Figure 11, we have made two adjustments: first, we used a full-page landscape layout to display the visualization results of the model detection effects; second, we locally enlarged four complex low-resolution scenes. The modified figure is now Figure 12 on page 18. Additionally, we have included explanations of the visualization results in lines 498-506 on page 17:

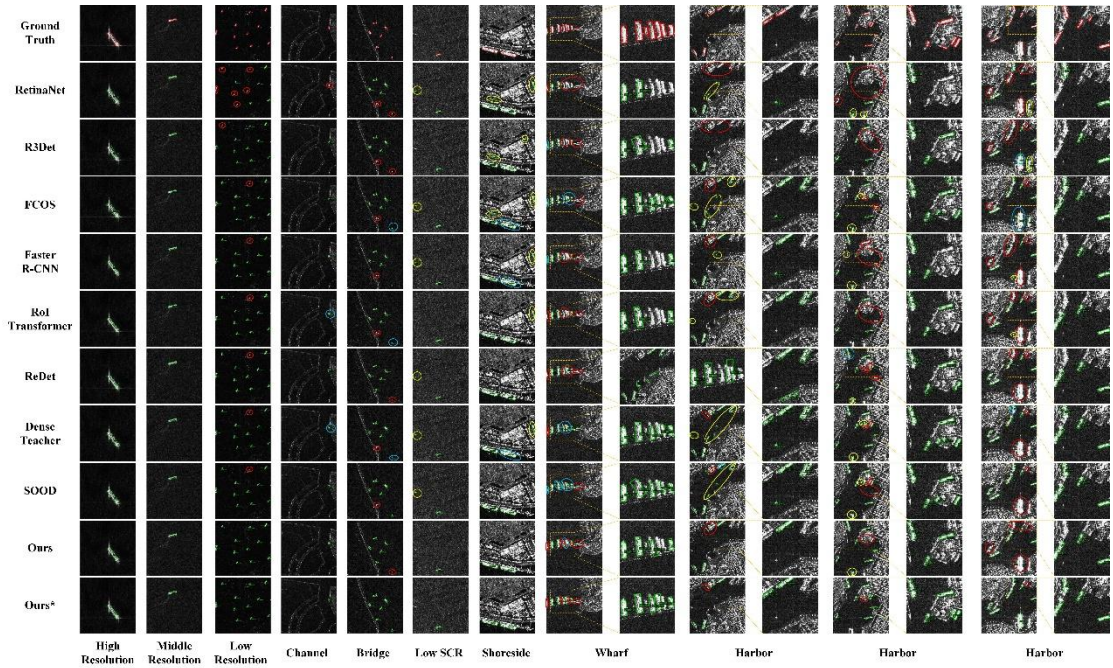


Figure 2. The visual comparison results of the algorithms mentioned in Table 4, where the red circle indicates missing detection, the yellow circle indicates falsealarm, and the blue circle indicates poor regression results of bounding box. The fewer and smaller the circles, the better the algorithm's performance. The images ofwharf and harbour are locally enlarged to achieve better visualization. * indicates that RDS is adopted.

" As shown in Figure 12, the scenes in the five images on the left are relatively simple, and the detection performance of the other methods, except for RetinaNet, is relatively good. In low SCR scenarios, most algorithms mistakenly detect interference on the left side of the image as ships, resulting in false alarms. The scenes in the five images on the right are more complex. The proposed method achieves the best detection performance, as indicated by the size and number of circles in the original images. The locally enlarged images also reveal more detailed detection results. Although the proposed method still has some missing detections, it is noteworthy that these results are obtained using only 10% training images. "

Major review:

Q1: However, the methodology is not clear. There are many secondary and tertiary procedures described. It makes the text understanding quite confusing. It is not clear where is the real scientific contribution. The authors made a technical work. However, the claimed scientific contribution seems to be weakly relevant.Thus, the present work needs to be rewritten. The methodology must be clear and the Authors must emphasize the scientific contribution. Particularly, I consider that a bunch of procedures can solve a problem. However, it is not a guarantee that is a scientific study. The Authors must emphasize clearly the relation cause-effect for each step of the proposed methodology.

Thank you very much for your valuable feedback. We have made the following modifications: In the original version, the 'Method' section only briefly introduced the implementation steps of each method without systematically explaining their relationships, which I believe may cause reading difficulties for other readers as well. At the beginning of Section 3, we have added an overall overview to further explain the issues this paper aims to solve, the methods used, the relationships between these methods, the relationship between these methods and the problem, and the specific operations in practical applications., including a very intuitive schematic diagram (Figure 3 on page 5) and an explanatory paragraph (lines 178-194 on page 5-6 and lines 195-201 on page 7) to help readers more easily and quickly understand the research approach of this paper. Additionally, we have created Table 1 on page 6 in the revised manuscript to describe the components of each method and briefly explain the roles of these methods.

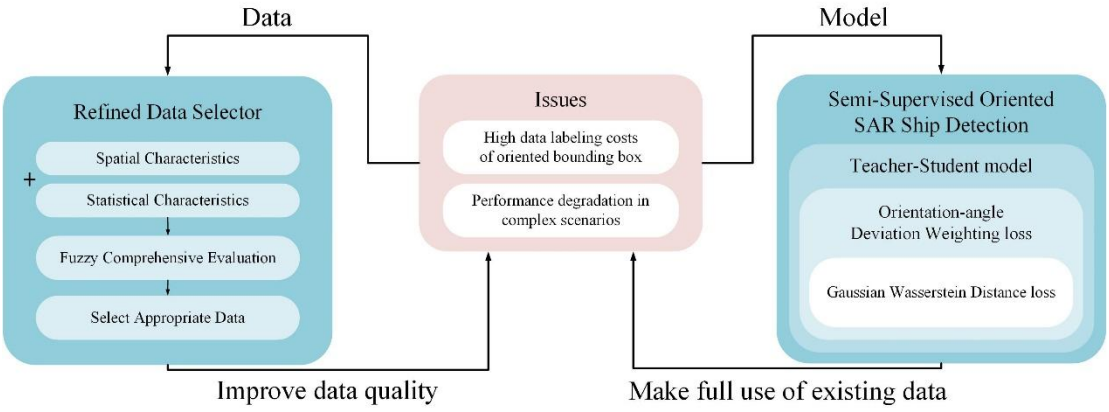


Figure 3. Overall structure of this paper.

" The research approach and methodology of this paper are shown as in Figure 3. To address the issues of high data labeling costs and performance degradation in complex scenarios for SAR ship detection, this paper proposes solutions from both data and model perspectives.

On the data side, a certain percentage of training data is selected as labeled data through RDS, which has a good representation and includes as many scenarios as possible, to enhance the model's training performance. First, we obtain the spatial and statistical features of SAR images as evaluation indicators, then use fuzzy comprehensive evaluation to score the images comprehensively, and finally select the appropriate data based on the comprehensive scores.

On the model side, we design a semi-supervised oriented SAR ship detection

framework that can fully utilize the existing labeled data and a large amount of unlabeled data. The specific process is shown as in Figure 2. For the first few iterations, known as the "burn in" stage, supervised training of the student model is performed using only labeled data, and the teacher model is updated by exponential moving average (EMA). In the unsupervised training stage, the unsupervised loss is calculated between pseudo-label prediction pairs."

Table 1 Descriptions and functions of different componets.

Component	Description	Function
RDS	Select appropriate indicators to evaluate SAR images, use FCE for comprehensive assessment, and filter data based on the final scores.	Obtain scores for SAR images and acquire a higher quality SAR dataset.
Spatial Characteristics	Obtained from the number, area, and spacing of connected regions after binarization, dilation, and other morphological operations.	Describing the spatial distribution of high-intensity pixels in SAR images, and used as evaluation indicators for FCE.
Statistical Characteristics	Including the mean and variance of SAR image gray-scale values, as well as some features of the histogram.	Describing the statistical distribution of SAR image pixels, and used as evaluation indicators for FCE.
FCE	Membership functions are derived from the distribution of evaluation indicators, and single factor evaluation is performed for different indicators. The final score is calculated using the weighted average fuzzy product.	Obtain comprehensive scores for SAR images for data selection purposes.
Select Appropriate Data	After obtaining the final scores, data selection is performed through interval sampling.	As the name implies.
Semi-Supervised Oriented SAR Ship Detection	A teacher-student model that combines supervised and unsupervised learning, using ODW loss as the unsupervised learning loss function.	Make full use of existing labeled data and leverage a large amount of unlabeled data to improve generalization ability
A teacher-student model that combines	During the unsupervised learning phase, only the student model is trained, and the teacher model is updated using EMA at the end of each iteration.	It allows for end-to-end semi-supervised learning.

ODW loss	The deviation between the student model's predictions and the teacher model's generated pseudo-labels is used as weights to dynamically weight the unsupervised training loss.	This improves the accuracy of the model's bounding box angle regression.
GWD loss	The OBB is modeled as a two-dimensional Gaussian distribution, and the Wasserstein distance between the student model's predictions and the pseudo-labels, as well as the ground truth, is calculated as the bounding box regression loss function.	As part of the ODW loss, this approach enhances the accuracy of the model's bounding box predictions, especially in low SCR and high sidelobe effect scenarios.

"The function of each component is shown as in Table 1. In practical applications, if a new SAR ship detection dataset needs to be established, the required amount of data is first selected using the Refined Data Selector (RDS), ensuring that the scenes are highly representative. These selected images are then labeled. Subsequently, the proposed semi-supervised learning model is trained using this small amount of labeled data along with the remaining large amount of unlabeled data. This approach ensures the model's detection performance while minimizing the labeling burden."

Q2: Moreover, the Authors must dedicate a Section describing in detail the dataset. In this current version, such a Section is mixed with experiments.

Thank you very much for your valuable feedback. We have split the original 'Datasets Description and Implementation Details' section into two separate sections: 'Datasets Description' and 'Implementation Details.' And we have made the following modifications: In the previous version of the manuscript, only the RSDD-SAR dataset was described in detail, while the SSDD, HRSID, and LS-SSDD datasets were not thoroughly explained. In this revised version, we have separated the dataset descriptions and the algorithm implementation details into two subsections and added explanations for the datasets in lines 374-386 on page 13.

"The Rotated Ship Detection Dataset in SAR images (RSDD-SAR) dataset mainly used in this paper is prepared for oriented ship detection, containing a training set of 5000 images and a test set of 2000 images, which are taken from 84 Gaofen-3 data, 41 TerraSAR-X data and 2 original large images, totaling 127 scenes. The images in the dataset contain different latitude and longitude, acquisition time, imaging mode, resolution, polarimetric mode, incidence angle, and imaging width. The oriented bounding box definition method of RSDD-SAR dataset is Long Edge definition, as shown in Figure \ref{def}, and the unit of angles is radians. Because the RSDD-SAR dataset has more diverse sources, we use it as the primary dataset for semi-supervised learning.

Three additional datasets are also used in this study: SAR Ship Detection Dataset (SSDD), High-Resolution SAR Images Dataset (HRSID) and Large-Scale SAR Ship Detection Dataset (LS-SSDD). SSDD is the first open dataset that is widely used to research state-of-the-art technology of DL-based SAR ship detection. In SSDD, there are typical hard-detected samples that need special consideration in practical SAR ship detection applications, such as small ships with inconspicuous features, densely parallel

ships berthing at ports with overlapping hulls and ships with large scale differences. HRSID is the first SAR ship dataset that supports instance segmentation. It has richer data sources and scenes compared to SSDD, but it uses only high-resolution SAR images. LS-SSDD contains 15 large-scale SAR images, accurately labeled with the aid of the automatic identification system (AIS) and Google Earth. "

Q3: At last, I suggest that the Authors search for related works that focus on "annotation". It seems that it could have good simpler and effective solutions. Doing a comparison between them and yours will be very useful.

Thank you very much for your valuable feedback. We have made the following modifications: We reviewed some literature on annotation and subsequently revised the "semi-supervised object detection" subsection in the Related Works section to "Addressing Labeling Costs" to encompass a broader scope, from lines 137 on page 4 to lines 176 on page 5. In addition to the original references on semi-supervised object detection, we also included content related to active learning in lines 159-167 on page 5. To maintain consistency in terminology throughout the text, we use "label" instead of "annotation". We compared our approach with active learning methods in lines 168-176 on page 5:

"*Addressing Labeling Costs*

……

In addition, weakly supervised learning, few-shot learning, and active learning also help address labeling costs. Unlike other methods that alleviate the problem from an algorithmic perspective, active learning addresses the issue from a data perspective. Active learning aims to maximize the training performance of the model while minimizing the amount of labeling required. In other words, it seeks to select the fewest but most useful samples from the unlabeled data for experts to label, thereby reducing labeling costs while maintaining training effectiveness. Active learning methods are also applied to SAR image recognition, and the results, after further labeling, can also serve as datasets for object detection.

Using model training methods for SAR data selection may be overly complex and unnecessary. Optical images possess rich texture details and colors, making it difficult to describe an optical image using simple indicators such as mean and variance. In contrast, SAR images typically appear as gray-scale images with fewer texture details, allowing for data selection using non-learning and simpler methods. In SAR ship detection tasks, considering the gray-scale differences between ships and the sea surface, the darker areas usually represent the sea surface, while the brighter areas represent ships, interference, or land clutter. Therefore, the statistical and spatial distributions of gray-scale values can effectively describe the scene and exhibit good generalization ability. "

Q4: I would like to suggest to the Authors that each paragraph can be written with a clear purpose, which means, it would have a subject discussed with the reasoning well established. Many times, along current version of the manuscript, one can be confused with the lack of idea connections among them.

Thank you very much for your valuable feedback. We have made the following modifications:

In the Related Works section, from line 82 on page 3 to line 176 on page 5, we reorganized the literature review. Since the paragraphs were lengthy, we started a new paragraph for the overall introduction and highlighted the existing issues at the end. In the Method section, we added simple, general explanatory statements at the beginning of some paragraphs or sections (while some already had such statements, or their lengths were too short to warrant additional ones). In order to help readers quickly understand the corresponding parts, we added these statements. In addition, we have reorganized the content starting from Section 3.2, adding general explanatory statements in lines 315-324 on page 10.

"**2. Related Works**

At present, many achievements have been made in the field of SAR ship detection and addressing labeling costs: SAR ship detection mainly uses constant false alarm rate (CFAR) detection and DL-based object detection, and addressing labeling costs mainly uses semi-supervised object detection.

2.1. SAR Ship Detection

CFAR is a classic target detection algorithm in the field of radar. The CFAR detection method dynamically adjusts the detection threshold according to the statistical characteristics of radar echo data to keep the radar false alarm rate constant. This classical method is also widely used in the detection of ship targets in SAR images. This method largely depends on the statistical modeling of sea clutter and the parameter estimation of the chosen model. Commonly used distributions include Gamma distribution, Log-normal distribution, Weibull distribution, and K distribution.

There are many improvements to the CFAR, among which superpixel segmentation-based methods are mainstream: ……

However, the performance is poor in complex scenarios such as strong clutter, nearshore and multiple targets. Additionally, CFAR detection suffers from low computational efficiency and poor transferability.

DL has become the mainstream research direction for ship detection in SAR images. DL-based SAR ship detection methods can be categorized into anchor-based and anchor-free classes, depending on whether the network model uses preset anchor boxes. Anchor-based methods can further be divided into single-stage and two-stage methods, depending on whether the network model needs to generate proposed regions.

There are many DL-based methods for SAR ship detection. Since the size of anchor boxes needs to be manually adjusted or set based on clustering of the training dataset, these fixed anchor boxes are not always suitable for other datasets with different distributions …...

However, since deep learning is a data-driven detection method, it has high requirements for the dataset.

2.2. Addressing Labeling Costs

Semi-supervised learning combines supervised and unsupervised learning, making full use of a small amount of labeled data along with a large amount of unlabeled data. This approach can effectively improve the model's generalization ability across different scenarios.

The pseudo-label method is a mainstream approach in semi-supervised object detection. Pseudo-label methods predict the pseudo-labels for unlabeled images by using

a pretrained model and then jointly train the model with both labeled and unlabeled data after augmentation ……"

line 203 on page 7

"Selecting an appropriate dataset is crucial in semi-supervised object detection. When using a semi-supervised object detection method, selecting a suitable dataset can reduce the workload and cost of annotation on the one hand, and also improve the performance of the model on the other hand. In general, a SAR image with strong interference, strong clutter, and various scattering points must be more complex than the scene of a calm sea. Since the evaluation of the complexity of SAR images from multiple aspects is fuzzy and subjective, fuzzy comprehensive evaluation (FCE) will make the results as objective as possible and conform to human's subjective feelings. "

line 220 on page 7

"Selecting appropriate evaluation indicators in FCE is the most fundamental step. Two SAR images and their gray-scale histograms are shown as in Figure 5, one is a complex scene A with strong interference, and the other is an offshore scene B with a calm sea. Obviously, the mean value and variance of Figure 5(a) will be significantly higher than that of Figure 5(b), and it is more complex spatially. Therefore, appropriate statistical characteristics of gray-scale and spatial characteristics can be selected as indicators of FCE…… "

line 238 on page 8

"The histogram analysis of SAR images reveals key characteristics for identifying ship targets and clutter. In the same SAR image, the gray-scale of the ship target is relatively strong and uniform, and the proportion of pixels is small, so it should be displayed as a peak with a higher gray-scale, narrower width and lower height on the histogram. Therefore, the more peaks there are, the wider the width of the widest peak, indicating that there may be strong sea clutter, interference or land clutter in the SAR image.…… "

lines 268-272 on pages 9

"3.1.2. Fuzzy Comprehensive Evaluation

The basic process of FCE is as follows: first, determine the evaluation set and the weights of evaluation indicators for SAR images, ranging from simple to complex. Next, conduct corresponding fuzzy evaluations for each indicator and determine the membership function. Then, form the fuzzy judgment matrix. Finally, perform fuzzy operations with the weight matrix to obtain a quantitative comprehensive evaluation result."

line 302 on page 10

"Appropriate training data is selected by interval sampling according to a certain *CS* distribution. According to the *CS* distribution of all data, the gray-scale values are manually divided into several intervals, and a certain amount of data is randomly

sampled in different intervals to obtain labeled training data with different *CS* distributions."

lines 314-323 on page 10

"3.2. Orientation-angle Deviation Weighting Loss

We design orientation deviation weighting (ODW) loss as the unsupervised loss to enhance the performance of semi-supervised learning. In supervised learning, the ground truth is used as a reliable reference, and the prediction results will be forced to move closer to it. However, in semi-supervised learning, we cannot simply take the pseudo-labels, generated by the teacher model, as the ground truth, and copying the supervised training will cause the effect of semi-supervised learning to deteriorate in a positive feedback style: the student model learns the wrong information from the unreliable pseudo-labels generated by the teacher model, and the EMA updated teacher model continuously gets the wrong "cognition" and generates more unreliable pseudo-labels, eventually leading to the deterioration of the performance of semi-supervised training."

lines 336-338 on page 11

"Considering that the difference in orientation-angle between pseudo-label prediction pairs can reflect the difficulty of the sample to a certain extent, this deviation can be used as a weight to adaptively adjust the unsupervised loss. All the unsupervised losses are dynamically weighted by the orientation-angle deviation of pseudo-label prediction pairs as the final unsupervised loss. The unsupervised bounding box losses are shown as in equation (12), and classification loss and centerness loss have similar forms. The supervised loss is also similar but without weight w_j."

And we have reorganized the presentation of the conclusions.

" In this paper, to reduce the labeling burden and improve the ship detection performance, we propose a semi-supervised oriented SAR ship detection framework from both data and model perspectives. We introduce a simple but effective scoring method based on FCE for SAR ship detection data. We also study the influence of the scoring distribution of labeled data on the training results of the semi-supervised model. RDS enhances the training effectiveness of the model by selecting more reasonably distributed data. The use of GWD loss and ODW loss improves the detection performance of the semi-supervised model in complex scenarios. The effectiveness of these proposed methods is validated through experiments. When creating a new SAR ship detection dataset creation, the RDS proposed in this paper can select appropriately distributed data for labeling. Subsequently, the semi-supervised model can be utilized for training. In the future, methods such as active learning and clustering can be used to further improve the quality of selected data."