Adjacent Teacher: Semi-Supervised Oriented Object Detection Leveraging Adjacent Spatial Consistency Prior in Remote Sensing Images

Tao Xia, Wei Jing, and Qi Wang, *Senior Member, IEEE*

***Abstract*—Oriented object detection in remote sensing images (RSIs) relies heavily on costly annotated data. To alleviate this challenge, we propose a straightforward yet powerful approach for semi-supervised oriented object detection, termed Adjacent Teacher. Drawing inspiration from the First Law of Geography, "Everything is related to everything else, but near things are more related than distant things". We observe that, in the adjacent space of RSIs, there is a widespread phenomenon that objects of the same category or closely related exhibit a clustered distribution and are roughly aligned in orientation. This discovery is generalized as the Adjacent Spatial Consistency Prior (ASCP), which reflects the consistent correlation between the categories and orientations of objects in the adjacent space of RSIs. Building on the ASCP, two novel modules are introduced: Low-Confidence Pseudo-Label Mining (LPM) and Pseudo-Label Angle Correcting (PAC). LPM boosts the number of reliable pseudo-labels by exploring low-confidence pseudo-labels that conform to the ASCP. PAC improves the quality of pseudo-labels by correcting their angles to satisfy the ASCP. With these, Adjacent Teacher achieves state-of-the-art results on the DOTA-v1.5, SODA-A and FAIR1M datasets, showing reduced missed detection rates and improved bounding box accuracy. Furthermore, the proposed method seamlessly integrates with existing pseudo-label-based semi-supervised oriented object detection models, significantly enhancing their performance. The code will be available at https://github.com/Xia-tao/Adjacent-Teacher.**

***Index Terms*—Remote sensing images, Oriented object detection, Semi-Supervised Learning, Adjacent Spatial Consistency Prior**

I. Introduction

With the rapid advancement of remote sensing technology, an extensive volume of imagery captured by satellites and unmanned aerial vehicles now provides a wealth of data resources for Earth observation. In this context, automated remote sensing image interpretation has become a prominent research field [1]–[9]. Object detection, a foundational task in automated interpretation, seeks to accurately identify and locate objects of interest within the images. Objects in remote sensing images (RSIs) frequently present arbitrary orientations, small scales, and clustered distributions. Such attributes make traditional horizontal bounding box detection methods unsuitable, primarily because of substantial redundancy in the bounding box and their limited effectiveness in distinguishing densely packed objects. To address these challenges, researchers have shifted towards oriented object detection, employing rotated bounding boxes to more precisely encapsulate and distinguish densely arranged objects in RSIs.

Fig. 1. Adjacent Spatial Consistency Prior. In the adjacent space of remote sensing images, there is a widespread phenomenon that objects of the same category or closely related exhibit a clustered distribution and are roughly aligned in orientation.

Oriented object detection involves predicting object orientations, offering a more comprehensive and precise approach for identifying objects with diverse poses in real-world scenarios. In recent years, remarkable progress has been made in this field by means of deep learning techniques. Researchers have improved the detection accuracy by angle representation [10]–[13], leveraging contextual information [2], and extracting rotation-invariant features [3], [14]. Despite these advancements, deep learning models require extensive annotations, which are labor-intensive and costly, limiting their broad applicability. For the oriented object detection tasks related to RSIs, the presence of small, densely distributed objects, coupled with the need for angle information, exacerbates the challenges associated with data annotation.

To mitigate the costs associated with data annotation, previous studies have often approached oriented object detection through the perspective of weakly supervised learning [15]–[19]. These researches have leveraged weak supervision signals, such as horizontal box annotations [15], [16] or even point annotations [17]–[19], to facilitate oriented object detection. To further diminish annotation costs, recent research has increasingly focused on semi-supervised object detection (SSOD). However, most SSOD methods [20]–[23] are primarily designed for objects in natural images, rendering them ill-

This work was supported by the National Natural Science Foundation of China under Grant U21B2041 and Grant 62471394. (Corresponding author: Qi Wang.)

Tao Xia is with the School of Computer Science, and with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, P.R. China. (e-mail: xiatao@mail.nwpu.edu.cn).

Wei Jing and Qi Wang are with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, P.R. China. (wei_adam@mail.nwpu.edu.cn; crabwq@gmail.com).

equipped to effectively detect dense, multi-scale, and arbitrarily oriented objects present in RSIs. This limitation arises from a failure to account for the geographical characteristics and directional information inherent in remote sensing contexts.

The First Law of Geography [24] states that "Everything is related to everything else, but near things are more related than distant things". The principle emphasizes the similarity and proximity between geographical phenomena and inspires our investigation into the spatial distribution patterns of objects in RSIs. As shown in Fig. 1, objects of the same category typically exhibit clustered distributions. Furthermore, the regularity of human activities is reflected in the structured arrangement of objects, as exemplified by the orderly parking of vehicles in the parking lot, sports fields methodically laid out in parks, and ships systematically moored in ports. Consequently, the distribution characteristics of objects in RSIs appear to adhere to the pattern whereby objects in close geographical proximity exhibit notable similarity and alignment. Specifically, in the adjacent space of RSIs, there is a widespread phenomenon that objects of the same category or closely related exhibit a clustered distribution and are roughly aligned in orientation. And we generalize the observation as the Adjacent Spatial Consistency Prior (ASCP).

In fully supervised tasks, the ASCP is implicitly included in the comprehensive annotations. However, in semi-supervised tasks, the prior information is disrupted by the incompleteness and inaccuracy of pseudo-labels, resulting in limited performance improvements. To tackle this problem, we propose Adjacent Teacher, a novel semi-supervised oriented object detection framework that leverages information from adjacent objects to guide pseudo-label selection and refinement. As illustrated in Fig. 2, Adjacent Teacher employs a teacher-student framework that incorporates both dense and sparse pseudo-labels. Additionally, it introduce two innovative modules: Low-Confidence Pseudo-Label Mining (LPM) and Pseudo-Label Angle Correcting (PAC). Unlike existing semi-supervised methods that treat pseudo-labels as isolated instances, Adjacent Teacher pioneers the use of ASCP to jointly optimize pseudo-label quantity (via LPM) and quality (via PAC). LPM uniquely mines reliable pseudo-labels from low-confidence objects by examining the ASCP with trusted neighbors, moving beyond conventional threshold-based paradigms. Meanwhile, PAC enhances the quality and accuracy of pseudo-labels by enforcing angle alignment with high-confidence adjacent objects. By integrating these modules, our approach not only enhances the quantity of reliable pseudo-labels while mitigating the risk of erroneous label incorporation, but also improves the angle accuracy, thereby significantly boosting the detection performance for oriented objects in semi-supervised learning scenarios. Overall, the main contributions of this paper are listed as follows:

1) Inspired by the First Law of Geography, we identify the prevalent Adjacent Spatial Consistency Prior (ASCP) in remote sensing images. Leveraging the ASCP, Adjacent Teacher is introduced as a method for semi-supervised oriented object detection.

2) Building upon the ASCP, we propose two key modules: the LPM module increases the number of reliable pseudo-labels by mining low-confidence pseudo-labels that conform to the ASCP; the PAC module improves the quality of pseudo-labels by correcting their angles to align with the ASCP.

3) Experimental results across multiple public datasets demonstrate that our method surpasses previous state-of-the-art approaches, exhibiting strong effectiveness and superiority. Furthermore, it can be seamlessly integrated into existing models, resulting in substantial performance improvements.

II. RELATED WORKS

Object detection based on deep learning is inherently data-driven, and the performance of models markedly relies on both the amount and quality of annotated data. However, obtaining precise object-level annotations in RSIs is labor-consuming, especially for oriented object detection, which requires the annotation of rotation angles. To address the challenge, researchers have investigated SSOD and oriented object detection under limited supervision.

A. Semi-Supervised Object Detection

Semi-Supervised Learning (SSL) is aimed at effectively making use of both labeled and unlabeled data during the training process. In recent years, substantial advancements have been achieved in image classification through SSL, yielding promising results. The existing SSL methods for image classification are mainly divided into two approaches: those based on pseudo-labels [25]–[27] and those based on consistency constraints [28]–[30]. Drawing upon these advances from semi-supervised image classification, similar strategies have been adapted for SSOD. For instance, in [31], detected boxes from unlabeled images are pasted onto labeled images, followed by position consistency estimation on the composite image. The CSD method [32] introduces left-right consistency constraints for object localization through horizontal flip data augmentation and proposes background erasure to mitigate the negative impact of background noise. However, its reliance on the relatively weak data augmentation of horizontal flips limits its overall performance. STAC [33] generates object-level pseudo-labels on unlabeled data in an offline manner, utilizes pre-trained detectors on labeled data, and updates the model using consistency constraints under strong data augmentation. Inspired by Mean Teacher [34], teacher-student networks have become widely adopted in SSOD [20]–[23], [28], [35]–[37]. In these networks, the teacher and student models have the same architecture but with different weights. The teacher model's weights are updated via the exponential moving average (EMA) of the student model's weights, while the student model is trained on the pseudo-labels produced by the teacher model. Unbiased Teacher [20] employs EMA and Focal Loss [38] to mitigate pseudo-label bias in SSOD, filtering out reliable pseudo-labels by using a fixed threshold. To better utilize on the teacher model's information, Humble Teacher [36] employs soft pseudo-labels for training the student model. Soft Teacher [21] adaptively weights the loss of unlabeled object according to the classification scores produced by the teacher network. and selects reliable pseudo-labels whose

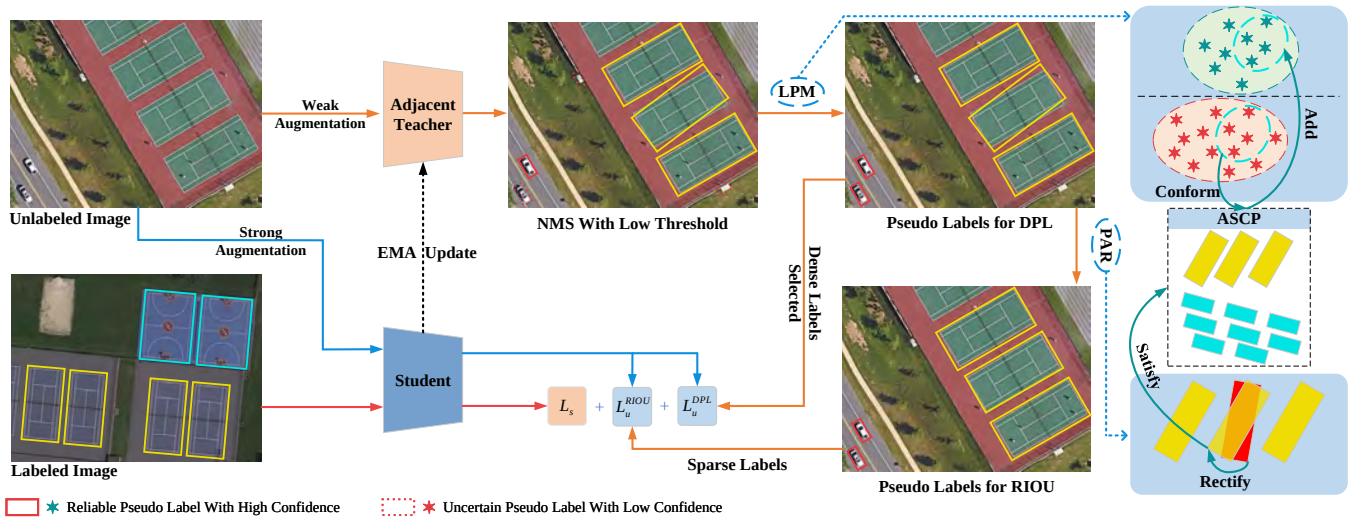


Fig. 2. The overview of the proposed Adjacent Teacher. The pseudo-labels of weakly augmented unlabeled images are generated by teacher model through NMS. LPM mine low-confidence pseudo-labels that conform to the ASCP. PAC rectify the angles of low-confidence pseudo-labels to satisfy the ASCP. The pseudo-labels derived from the LPM are employed to produce dense pseudo-labels, calculating $\mathcal{L}_u^{DPL}$. The pseudo-labels from both the LPM and PAC are used to calculate $\mathcal{L}_u^{RIOU}$. The overall unsupervised loss can be expressed as the weighted summation of the two losses.

box regression variance falls below a certain threshold. Unbiased Teacher v2 [23] firstly applies SSOD to an anchor-free detector (FCOS [39]), introducing the Listen2Student mechanism, which employs unsupervised regression loss and uncertainty prediction to eliminate misleading false labels during regression training of the bounding box. Dense Teacher [22] replaces object-level sparse predictions with pixel-level dense predictions and presents a method for selecting regions to highlight crucial details while reducing interference from dense labels. Consistent-Teacher [37] addresses the problem of unstable pseudo-labels causing noise during the training of student networks in SSOD, which can lead to severe overfitting. It proposes adaptive anchoring assignments to replace static IoU-based strategies, enabling student networks to better resist noisy pseudo-boundary boxes.

While these approaches can enhance performance in semi-supervised settings, they are not specifically designed for oriented object detection in RSIs. Consequently, they are unable to leverage the critical angle information and geographic attributes of oriented objects in RSIs.

B. Oriented Object Detection

In contrast to general object detection using horizontal bounding box (HBB), oriented object detection positions objects using oriented bounding boxes (OBB). Application scenarios for oriented object detection include object detection in RSIs [1], [14], [40] and text detection [41]–[43]. Depending on the angle prediction strategy, oriented object detection approaches can be classified as: angle regression [1], [14], [40], [41], [44], angle classification [10], [11], and Gaussian-based [12], [13] approaches.

RRPN [41] utilizes rotating anchors to obtain rotating Regions of Interest (ROI), enabling the extraction of features necessary for detecting scene text in arbitrary directions. RR-CNN [1] introduces the RRoI pooling layer to extract features from rotating objects, facilitating oriented ship detection. R3Det [40] enhances detection speed through progressive regression from predicted HBB to OBB. Based on horizontal anchors, ROI Transformer [14] derives rotating ROIs through a fully connected neural network in the Region Proposal Network (RPN) stage to reduce computation and employs RRoI Warping to extract rotation-invariant features. Oriented R-CNN [44] introduces an efficient rotated RPN network and employs midpoint offset to represent proposals in arbitrary directions. All of these methods rely on direct angle regression, which encounters the problem arises from boundary discontinuity caused by the periodic nature of angles. CSL [10] converts the angle regression problem into the classification task of angles, using Gray Code to encode angles and reduce encoding length. DCL [11] further refines angle coding with a more compact method, simplifying the prediction head and improving the prediction of square-like objects. GWD [12] converts the representation of oriented bounding boxes into a two-dimensional Gaussian distribution, using the Gaussian-Wasserstein-Distance as a regression loss to tackle problems related to boundary discontinuity and the non-differentiability of Rotation Intersection over Union (RIOU). Building on GWD, KLD [13] replaces the Gaussian-Wasserstein-Distance with Kullback-Leibler Divergence, enhancing the detection performance for objects with high aspect ratios.

In contrast to these fully supervised approaches, our work introduces a semi-supervised framework for oriented object detection. By leveraging both labeled and unlabeled data, our method significantly reduces the reliance on extensive high-quality annotations while maintaining competitive detection performance.

C. Oriented Object Detection Under Limited Supervision

The challenge of exorbitant annotation costs in oriented object detection has catalyzed the exploration and advancement of limited-supervised oriented object detection algorithms.

These algorithms include weakly supervised situations with abundant horizontal box annotations or point annotations, as well as semi-supervised situations where only a restricted number of annotated oriented bounding boxes are accessible.

Creating a minimum bounding rectangle around instance segmentation regions can yield oriented bounding boxes for objects. Therefore, weakly supervised instance segmentation with HBB provides an approximate solution. SDI [45] utilizes horizontal bounding box annotations to achieve instance segmentation by incorporating exploratory iterative training and refinement processes as a denoising method. BBTP [46] introduces a weakly supervised instance segmentation approach in accordance with the tightness prior of bounding boxes. BoxInst [47] leverages the prior information that neighboring pixels with comparable colors are probably part of the same category, designing a new loss function that converts pixel-level segmentation loss into object-level loss. However, the HBox-Mask-RBox workflow is overly complex, leading to reduced detection efficiency and accuracy. H2RBox [15] innovates with a more efficient workflow by eliminating the segmentation step, enabling direct learning of oriented object detection from HBB. WSODet [48] achieves oriented object detection by generating angle information from feature maps. H2RBox-v2 [16] further explores angle regression through symmetry supervision in oriented object detection, utilizing mirror symmetry and rotational consistency to learn object orientation. With less supervision using only point annotation information, PMHO [19] proposes a Point-Mask-HBox-Rbox workflow based on SAM. PointOBB [18] uses scaled-up views to help the network learn basic scale perception and employs rotation-flip views for angle prediction. Point2RBox [17] enhances box regression by sampling around each point and synthesizing the object, allowing the network to perceive object scale and angle through transformations in a self-supervised manner.

In contrast to the extensive research on oriented object detection with weak supervision, the field of semi-supervised oriented object detection remains nascent. SOOD [49] pioneered semi-supervised oriented object detection by establishing a robust baseline and serving as a foundation for further investigation. SOOD implements an adaptive weighted consistency regularization loss that relies on the angular discrepancy between teacher-student model predictions, while explicitly establishing a many-to-many correspondence between the pseudo-label set and the prediction set to enforce global consistency constraints. Different from SOOD, our research utilizes the Adjacent Spatial Consistency Prior in RSIs to augment both the number and accuracy of pseudo-labels, thereby enhancing the performance of all pseudo-label-based semi-supervised oriented object detectors.

III. Prior Verification

The Adjacent Spatial Consistency Prior (ASCP) hypothesizes that geographically proximate objects of the same category in remote sensing images generally exhibit the following two spatial patterns: spatial aggregation and angle consistency. We validate this hypothesis across three datasets: DOTA-v1.5, SODA-A and FAIR1M.

TABLE I
AVERAGE DENSITY AND ANGULAR CONSISTENCY BY CATEGORY FOR DIFFERENT DATASETS. THE **CLASS-IGNORE** REPRESENTS THE GLOBAL ANALYSIS ACROSS THE ENTIRE DATASET, EXCLUDING SPECIFIC CATEGORIES. - REPRESENTS THE ABSENCE OF THIS CATEGORY. **NONE** INDICATES UNIFORM ANGLE DISTRIBUTION OR NO SPATIAL AUTOCORRELATION.

Category	DOTA		SODA-A		FAIR1M	
	N_a	Moran's I	N_a	Moran's I	N_a	Moran's I
PL	0.63	0.39	0.34	0.36	0.14	0.24
BD	0.01	None	-	-	0.03	None
BR	0.14	**0.57**	-	-	0.21	**0.73**
GTF	0.04	None	-	-	-	-
SV	**1.36**	**0.69**	**1.24**	**0.79**	**0.85**	**0.57**
LV	**2.63**	**0.65**	**1.35**	**0.65**	**0.85**	**0.57**
SH	**2.19**	**0.56**	**2.01**	**0.62**	**1.92**	**0.65**
TC	**1.27**	0.34	-	-	**1.09**	0.42
BC	**1.03**	**0.79**	-	-	**0.69**	**0.54**
ST	0.19	0.23	0.31	0.50	-	-
SBF	0.35	None	-	-	0.24	None
RA	0.00	None	-	-	0.01	None
HA	**1.05**	**0.52**	-	-	-	-
SP	0.16	0.43	0.04	0.32	-	-
HC	**3.62**	0.22	0.19	0.31	-	-
CC	**1.89**	**0.56**	-	-	-	-
CO	-	-	**4.59**	**0.73**	-	-
WM	-	-	0.00	0.55	-	-
IS	-	-	-	-	0.00	None
Class-Ignore	**2.09**	**0.67**	**1.72**	**0.73**	**0.91**	**0.59**

A. Spatial Aggregation Measurement

We define a circular adjacent space with radius $r = \max(l, w)$ around each object's centroid, where l and w are its length and width. The number of same category objects within this adjacent space quantifies spatial aggregation. The average density N_a is calculated for each category. Since the adjacent space is defined only within an object's single-scale space, pronounced spatial aggregation occurs as N_a approaches 1. Table I summarizes N_a for each category. Categories like vehicles, ships, and containers indicate strong spatial aggregation ($N_a > 0.8$).

B. Angle Consistency Analysis

We use the Moran's I index [50] to measure the angle consistency. The Moran's I index measures spatial autocorrelation, ranging from -1 to 1, with higher values indicating greater similarity between adjacent regions. In this analysis, Moran's I was calculated for angle observations. Table I shows Moran's I for each category. Many categories, such as bridges, vehicles, basketball courts, and containers, exhibit strong angle consistency (Moran's I > 0.5).

Furthermore, statistical analysis reveals that objects exhibiting both spatial aggregation and angle consistency constitute 82.3%, 85.7%, and 88.7% of the total annotated instances in the DOTA-v1.5, SODA-A, and FAIR1M datasets, respectively. This confirms the prevalence of these patterns and supports the ASCP.

IV. PROPOSED METHOD

For semi-supervised object detection, the performance of the detector relies heavily on both the number and accuracy of pseudo-labels. Current pseudo-label selectors are primarily

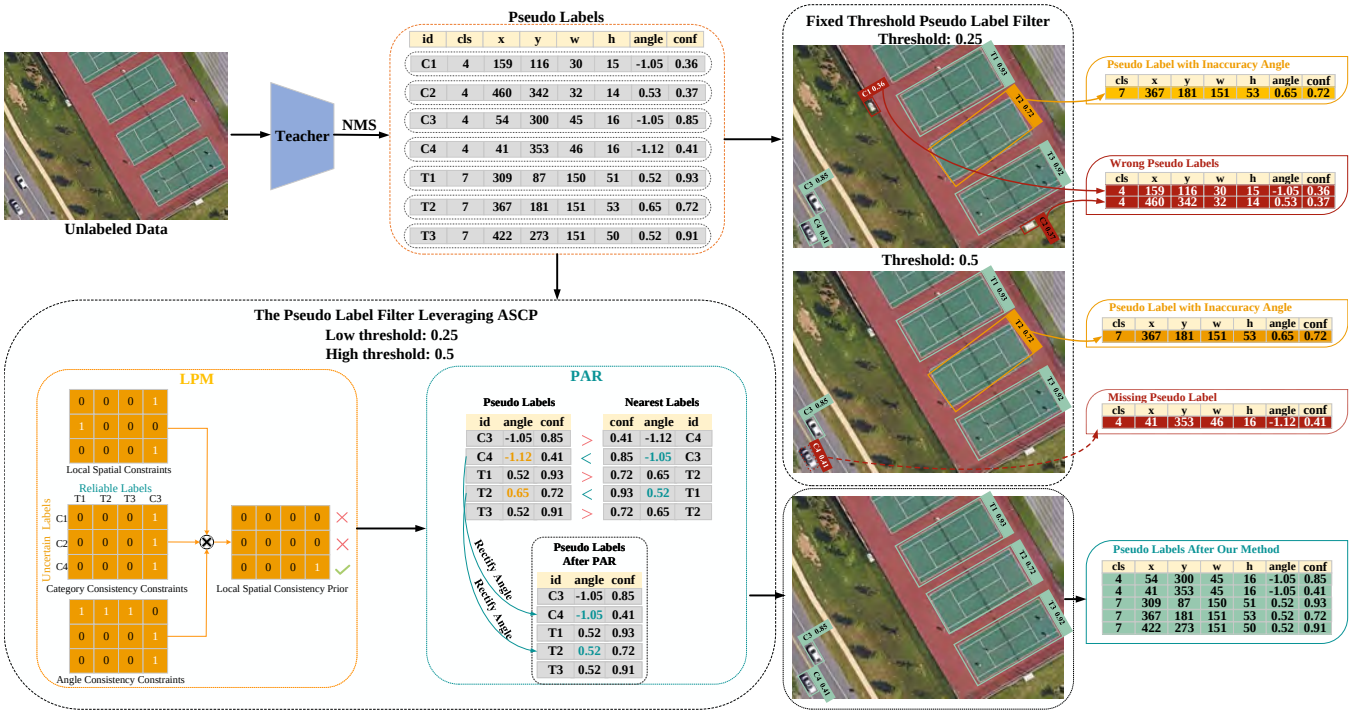


Fig. 3. Comparison of different pseudo-label filter: Fixed threshold pseudo-labels filter is prevalently employed in SSOD models, a low threshold (0.25) may lead to wrong pseudo-labels (indicated by by the bounding box with red solid line), while a high threshold (0.5) may miss correct pseudo-labels (indicated by the bounding box with red dashed line), and it fails to correct pseudo-labels deviating from the correct angle (indicated by the bounding box with solid yellow line). Our method is based on the ASCP, where LPM mines pseudo-labels that have low confidence but conform to the ASCP, PAC compares the confidence scores of pseudo-label pairs closest in distance, correcting the angles of low-confidence pseudo-labels to align with those of higher confidence, effectively addressing the aforementioned problems.

designed for traditional horizontal object detection and do not fully leverage the unique characteristics of objects in RSIs.

In the adjacent space of RSIs, it is common to observe that objects belonging to the same category or being closely related tend to cluster together and exhibit a roughly aligned orientation. We summarize the finding as the Adjacent Spatial Consistency Prior (ASCP) and validated it in three public RSIs datasets. Building on the ASCP, we introduce Adjacent Teacher, a semi-supervised oriented object detection method. In the proposed model, the Low-Confidence Pseudo-Label Mining (LPM) module increases the pool of reliable pseudo-labels by identifying low-confidence pseudo-labels that adhere to the ASCP. Meanwhile, the Pseudo-Label Angle Correcting (PAC) module enhances the quality of pseudo-label by correcting their angles to satisfy the ASCP.

In this section, we detail the proposed Adjacent Teacher, which is structured into three parts. First, section 3A introduces the overall architecture of the basic pseudo-label paradigm. Next, section 3B details the design of the LPM module. Finally, section 3C describes the PAC module.

A. The Basic Pseudo-Label Paradigm

The pseudo-label framework, which inherits the teacher-student model structure from Mean Teacher, has become a mainstream approach in SSOD. Given a set of fully labeled samples $I_l = \{x_i^l, g_i^l\}_{i=1}^{N_l}$ and a set of unlabeled samples $I_u = \{x_i^u\}_{i=1}^{N_u}$, x_i and g_i denote the image and the ground truth (GT) of i-th sampler, N_l and N_u denote the number of the I_l and I_u. During each training iteration, training samples are randomly selected from I_l and I_u according to the sampling rate. The student model acquires knowledge from both the labeled images having GT and the unlabeled images with pseudo-labels produced by the teacher model, and updates the model through gradient back-propagation. Thus, the overall loss function of the student model is expressed as the weighted sum of the supervised loss $\mathcal{L}_{sup}$ for labeled images and the unsupervised loss $\mathcal{L}_{unsup}$ for unlabeled images:

$$\mathcal{L} = \frac{1}{N_l}\sum_{i=1}^{N_l}\mathcal{L}_{sup}(x_i^l, g_i^l) + \frac{1}{N_u}\sum_{i=1}^{N_u}\alpha\mathcal{L}_{unsup}(x_i^u, \widehat{g_i^u}), \quad (1)$$

where $\widehat{g_i^u}$ denotes the pseudo-labels of i-th unlabeled image generated by the teacher model and α controls contribution of unsupervised loss. During training progress, the teacher model is continuously updated based on the EMA of the student model.

FCOS [39] is chosen as the base detector. Drawing on insights from Unbiased Teacher v2 [23], it was discovered that center-sampling strategy, as introduced in FCOS, does not appear to provide significant assistance and may even impede oriented object detection in semi-supervised settings. Therefore, we remove the center-sampling strategy from our model to avoid any negative impact on performance.

We adopt the Dense Pseudo-Label (DPL) framework from SOOD [49]. For the bounding box regression, the L1 loss is

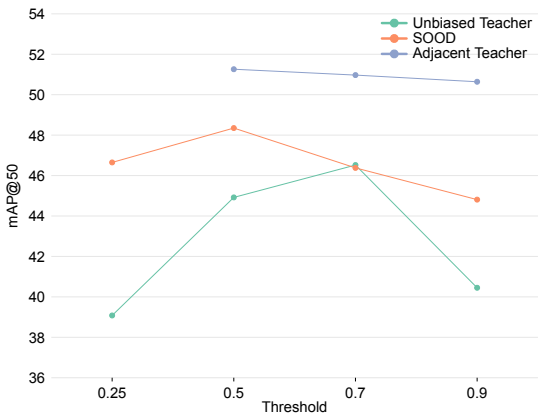


Fig. 4. The impact of different fixed thresholds for the performance of SSOD models using fixed pseudo-label filter.

employed, and for classification and centerness, binary cross-entropy loss is used. The $\mathcal{L}_{DPL}$ is formalized as:

$$\mathcal{L}_{DPL} = w^{rot}(\mathcal{L}_{cls} + \mathcal{L}_{center} + \mathcal{L}_{bbox}), \tag{2}$$

$$w^{rot} = \beta \frac{|r^t - r^s|}{\pi}, \tag{3}$$

where r^t and r^s are the angle predictions of the teacher model and student model, respectively, in radians. β is a hyper-parameter for adjusting the importance of orientation, and SOOD [49] sets it to 50.

However, correcting the orientation of dense pseudo-labels presents a challenge. To address it, the Rotated Intersection over Union (RIOU) loss is employed to quantify the disparity between the sparsely corrected pseudo-labels and the predictions of the student model, serving as a component of the unsupervised loss. Consequently, the unsupervised loss is determined as the weighted sum of $\mathcal{L}_{DPL}$ and $\mathcal{L}_{RIOU}$:

$$\mathcal{L}_{unsup} = \mathcal{L}_{DPL} + \gamma \mathcal{L}_{RIOU}, \tag{4}$$

where γ controls contribution of RIOU loss and is empirically set to 0.5.

B. Low-Confidence Pseudo-Labels Mining

Assigning pseudo-labels is a major challenge in SSOD. Previous studies [20], [21], [49] employed a fixed threshold τ to assign pseudo-labels, considering those with confidence scores greater than τ as objects and the others as background. As illustrated in Fig. 3, setting a low threshold introduces a large number of wrong pseudo-labels, while a high threshold results in the omission of many correct pseudo-labels. Fig. 4 shows that both scenarios introduce significant noise into the training process, leading to severe over-fitting issues. LPM addresses the specific challenge by introducing a low-confidence pseudo-label mining method based on the ASCP, with strong robustness to threshold selection. The process of LPM is as detailed in Algorithm 1.

Algorithm 1 Low-Confidence Pseudo-Labels Mining

Input: pseudo-label set P from the teacher model after NMS, threshold τ_g for grouping, parameter λ for adjacent spatial constraints, map function M for category consistency constraints, threshold τ_{ang} for angle consistency constraints, an empty set E
Output: reliable pseudo-label set H

```
Get high-confidence pseudo-label set L by Eq.(3)
Get high-confidence pseudo-label set H by Eq.(4)
Get adjacent spatial constraints matrix S by Eq.(7)
Get category consistency matrix C by Eq.(4)
Get angle consistency matrix A by Eq.(9)
Adjacent spatial consistency prior matrix P ← S × C × A
for j ← 0 to |L| − 1 do
    for i ← 0 to |H| − 1 do
        if P_{i,j} = 1 then
            Add L_j to E
            continue
        end
    end
end
H ← H ∪ E
return H
```

We assign the pseudo-label set P obtained after non-maximum suppression (NMS) by the teacher model into two groups based on their confidence scores:

$$L = \{l_i \,|\, \frac{\tau_g}{2} <= l_i^{conf} < \tau_g, \, l_i \in P\}, \tag{5}$$

$$H = \{h_i \,|\, h_i^{conf} >= \tau_g, \, h_i \in P\}, \tag{6}$$

where L refers to the low-confidence pseudo-label group, treated as uncertain pseudo-labels, and H refers to the high-confidence pseudo-label group, treated as reliable pseudo-labels. τ_g is the threshold used for grouping. Subsequently, the pseudo-labels in H are used as a reference to identify and extract the pseudo-labels in L that adhere to the ASCP. The process involves the following three constraints:

1) Adjacent Spatial Constraints: The adjacent space is defined as the area in which the distance between two objects is less than the specified distance threshold.

As illustrated in Fig. 5a, the presence of large-scale objects in RSIs renders the Euclidean distance between the center points of two objects (depicted by the orange line) an unreliable measure of spatial adjacency. Furthermore, as shown in Fig. 5b, due to the significant scale variations among objects in RSIs, it is inappropriate to employ a uniform distance threshold to all objects. For example, the threshold for determining adjacency between tennis courts, owing to its larger scale, should be greater than that applied to cars. Consequently, a method for calculating an object-scale-dependent adaptive distance threshold is proposed as follows:

$$\boldsymbol{D}_{(i,j)} = \lambda * \min(\max(h_i^w, \, h_i^h), \, \max(l_j^w, \, l_j^h)), \tag{7}$$

where h_i^w and h_i^h denote the width and height of the i-th object in H, l_i^w and l_i^h denote the width and height of the j-th object in L. The parameter λ controls the size of the adjacent space

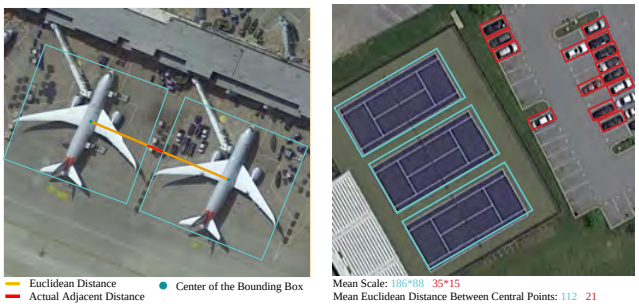


Fig. 5. (a) the proximity between large scaled objects cannot be measured by the Euclidean distance between their center points; (b) the scale differences among distinct classes of objects result in significant variations in the average Euclidean distances between each classes, making it inappropriate to apply a uniform distance threshold for all objects when considering spatial adjacency.

definition. Finally, the adjacent spatial constraints matrix S can be formalized as:

$$\boldsymbol{S}_{(i,j)} = \begin{cases} 1, & \text{if } \mathrm{dis}(h_i, l_j) <= \boldsymbol{D}_{(i,j)}, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where dis(,) denotes the function for calculating the Euclidean distance between the center points of two objects.

2) Category Consistency Constraints: Not only objects within the same category, but also closely related object categories should adhere to the category consistency constraint, such as small cars and large cars, or ships and harbors. Therefore, a function M is defined to map closely related categories to the same new category, and the category consistency matrix $\boldsymbol{C}$ can be formulated as:

$$\boldsymbol{C}_{(i,j)} = \begin{cases} 1, & \text{if } \mathrm{M}(h_i^{cls}) = \mathrm{M}(l_j^{cls}), \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

where h_i^{cls} denotes the category of the i-th object in H, l_j^{cls} denotes the category of the j-th object in L, $\boldsymbol{C}_{(i,j)}$ represents whether the i-th object in H and the j-th object in L conform to the category consistency constraint.

3) Angle Consistency Constraints: The angle consistency constraint is defined as the requirement that the angle difference between two objects must be less than a threshold τ_{ang}. As illustrated in Fig. 6a, we follow a common angle representation method. However, this approach for angle definition will lead to boundary problem. As shown in Fig. 6b, when the angles of two objects approach $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ respectively, the computed angle difference can be significant, despite the angles appearing visually quite similar. To address it, the case that the angle difference approaches π is also regarded as satisfying the angle consistency constraint. Finally, the method for calculating the angle consistency matrix $\boldsymbol{A}$ can be formalized as:

$$\boldsymbol{A}_{(i,j)} = \begin{cases} 1, & \text{if } |h_i^{ang} - l_j^{ang}| < \tau_{ang}, \\ 1, & \text{if } |h_i^{ang} - l_j^{ang}| > \pi - \tau_{ang}, \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

where h_i^{ang} denotes the angle of the i-th object in H, l_j^{ang} denotes the angle of the j-th object in L, and $\boldsymbol{A}_{(i,j)}$ represents

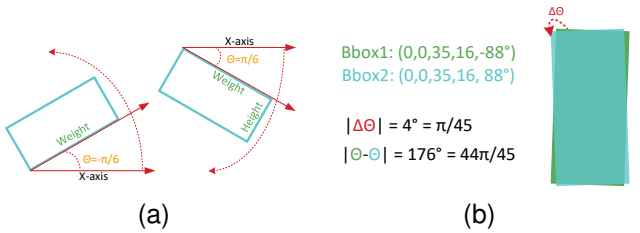


Fig. 6. (a) The angle θ is defined as the angle between the X-axis and the longer side of the bounding box, constrained within the range $[-\frac{\pi}{2}, \frac{\pi}{2})$; (b) the boundary issue stemming from the representation of angles.

whether the i-th object in H and the j-th object in L conform to the angle consistency constraint. The parameter τ_{ang} controls the strictness of angle consistency.

When the object pairs simultaneously satisfy all three constraints, they are considered to conform to the ASCP. Therefore, the adjacent spatial consistency prior matrix $\boldsymbol{P}$ is expressed as $\boldsymbol{P} = \boldsymbol{S} \times \boldsymbol{C} \times \boldsymbol{A}$. In the low confidence pseudo-label group L, the object set E that conforms to the ASCP can be formalized as:

$$E = \{l_j \mid \boldsymbol{P}_{i,j} = 1,\ l_j \in L\}. \tag{11}$$

Since the object in E exhibit strong consistency with the reliable objects in H, the objects in E are also considered to be reliable. Therefore, the set H is updated as $H = H \cup E$.

C. Pseudo-Label Angle Correcting

The regularity of human activities is prominently reflected in the ordered arrangement of objects in RSIs. The orderliness is evident in the spatial arrangements of adjacent objects, making angle consistency an important characteristic. As illustrated in Fig. 3, prior studies have overlooked this critical pattern, leading to inaccuracies in the angles of pseudo-labels (indicated by solid yellow lines). To address it, PAC employs the ASCP to rectify the angles of low-confidence pseudo-labels.

Algorithm 2 Pseudo-Label Angle Correcting

Input: pseudo-label set H
Output: refined pseudo-label set H.
foreach *pseudo-label* h_i *in* H **do**
  Find its nearest object within the same category h_j
  if $h_i^{conf} < h_j^{conf}$ **then**
    $h_i^{\text{ang}} \leftarrow h_j^{\text{ang}}$
  end
end
return H

Algorithm 2 describes how the proposed method works. Firstly, a reasonable assumption is established that higher confidence scores typically correspond to more accurate angle predictions. Next, the adjacent space of each object is defined as its nearest neighbor within the same category. Finally, by comparing the confidence of this pair of pseudo-labels, the angle of the lower-confidence pseudo-label is corrected to match angle of the higher-confidence one. PAC refines the angle accuracy of lower-confidence pseudo-labels in H, thus enhancing the accuracy of pseudo-labels.

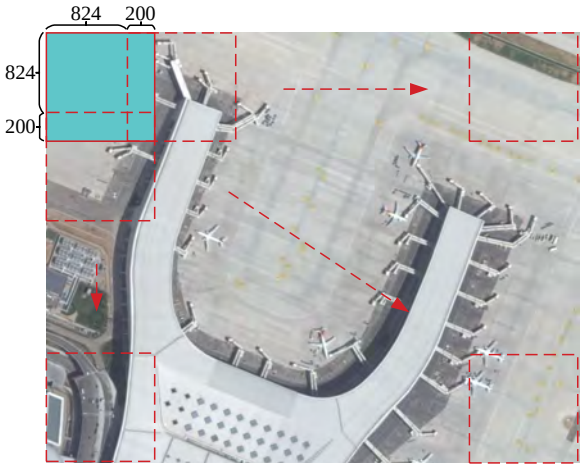


Fig. 7. The image cropping strategy is implemented where the original images are cropped into 1024 × 1024 pixel patches with a stride of 824 pixels, ensuring a 200 pixels overlap between consecutive patches.

V. EXPERIMENTS

A. Datasets and Evaluation Metrics

1) DOTA-v1.5: DOTA-v1.5 [51] is a large-scale RSIs dataset comprising a diverse array of objects with varying orientations, scales, and appearances. DOTA-v1.5 uses OBB to annotate 40,289 objects across 16 categories: plane (PL), ship (SH), harbor (HA), baseball diamond (BD), bridge (BR), swimming pool (SP), helicopter (HC), small vehicle (SV), large vehicle (LV), ground track field (GTF), tennis court (TC), basketball court (BC), soccer ball field (SBF), storage tank (ST), roundabout (RA), and container crane (CC).

2) SODA-A: The images in the SODA-A dataset [52] are derived from a multitude of cities around the world, showcasing a rich tapestry of data diversity. The number of instances per image ranges from 1 to 11,134, covering both sparse and dense scenes. Instances in SODA-A are categorized into 9 classes: PL, HC, container (CO), windmill (WM), ST, SH, SP, SV, and LV.

3) FAIR1M: The FAIR1M dataset [53] includes more than than 15,000 remote sensing images and more than 1 million instances, with image resolutions ranging from 0.3m to 0.8m. The dataset contains objects spanning 5 major categories and 37 subcategories. In this experiment, we focus on the three major categories of vehicles (SV, LV), SH, and PL, as well as seven subcategories: SBF, BC, TC, BD, RA, BR, and intersection (IS).

4) Metrics: In our experiments, the evaluation metric to assess the accuracy of detected objects in the validation set is the standard mean average precision (mAP) at IOU thresholds of 0.5 and 0.75.

B. Implementation Details

As shown in Fig. 7, large-scale RSIs are cropped into 1024×1024 patches (824-pixel stride, 200-pixel overlap). To promote consistency in predictions for perturbed data, weak (random resizing/flipping) and strong (color/contrast/sharpness adjustments, grayscale conversion, Gaussian blurring) augmentations are applied to unlabeled data.

TABLE II
SUMMARY OF HYPER-PARAMETERS. {,} REPRESENTS GROUPING CATEGORIES WITHIN THE SET INTO A NEW CATEGORY.

Hyper-parameter	Value	Reason for Selection
τ_g	0.5	Experiments (Fig. 4)
λ	2.0	Experiments (Table X)
M	{SV, LV}, {TC, BC}, {SH, HB}.	Experiments (Table XI)
τ_{ang}	0.15	Experiments (Table XII)

We take the oriented version of FCOS [39] as the detector and adopt the RAW loss from SOOD [49]. The models are implemented under MMRotate [54] framework. The model undergoes training for 120,000 iterations with a batch size of 6 images, maintaining a 2:1 ratio between labeled and unlabeled images, operating on a single RTX 3090 GPU. Employing the stochastic gradient descent (SGD) optimizer, the initial learning rate is set to 0.0025, and is adjusted by a factor of 10 at the 80,000 and 100,000 iteration milestones. The parameters of momentum and weight decay are respectively assigned values of 0.9 and 0.0001. And the "burnin" strategy is employed for initializing the teacher model. To ensure reproducibility, the random seed was fixed at 42, and all reported results represent the average of three independent trials.

As the proposed method involves multiple hyper-parameters, Table II provides a detailed list of these hyper-parameters to promote the reproducibility of the method. The process of selecting the optimal hyper-parameters begins by choosing a set of reasonable values based on the practical significance of each parameter, followed by conducting experiments on the DOTA-v1.5 dataset to select the corresponding optimal hyper-parameter values based on the best experimental results.

In these experiments, two protocols are considered to validate the effectiveness of our approach in semi-supervised tasks: data with partial labeled and data with fully labeled.

In the experiments with partially labeled data, a certain proportion of images in the training set are sampled as labeled dataset. The remainder of the training data is regarded as an unlabeled dataset. To ensure fairness in comparison, the partitioning of the DOTA-v1.5 dataset strictly follows SOOD [49]. The SODA-A and FAIR1M datasets are randomly sampled, ensuring that the partially labeled set maintains a similar data distribution to that of the entire training dataset.

In the experiments with fully labeled data, the whole data for training is treated as a labeled dataset, while the data for testing, which lacks released annotations, is regarded as unlabeled dataset.

C. Compare With the State-of-the-art Methods

Our Adjacent Teacher is compared with state-of-the-art (SOTA) models on two datasets, including Unbiased Teacher [20], Soft Teacher [21], Dense Teacher [22], and SOOD [49]. Since we focus on oriented objects in remote sensing scenes, all detectors are specifically implemented to detect oriented objects. In the table, the symbols $*$ and † represent the

TABLE III
QUANTITATIVE RESULTS ON DOTA-V1.5 DATASET UNDER THE CONFIGURATION OF PARTIALLY LABELED DATA.

Setting	Method	Publication	10%		20%		30%	
-	-	-	mAP_{50}	mAP_{75}	mAP_{50}	mAP_{75}	mAP_{50}	mAP_{75}
Supervised	Faster R-CNN* [55]	TPAMI 2016	42.13	-	48.89	-	53.14	-
	FCOS† [39]	ICCV 2019	44.63	-	52.41	-	54.77	-
Semi-supervised	Unbiased Teacher* [20]	ICLR 2021	46.39	23.48	53.04	26.99	53.56	27.64
	Soft Teacher* [21]	ICCV 2021	49.95	21.83	57.05	27.43	59.53	29.99
	Dense Teacher† [22]	ECCV 2022	48.87	23.26	55.22	27.49	59.30	28.70
	SOOD† [49]	CVPR 2023	48.63	23.10	55.58	28.43	59.23	29.55
	Adjacent Teacher†(**ours**)	-	**51.26**	**24.09**	**57.42**	**29.16**	**60.26**	**30.68**

TABLE IV
QUANTITATIVE RESULTS ON SODA-A DATASET UNDER THE CONFIGURATION OF PARTIALLY LABELED DATA.

Setting	Method	Publication	10%		20%		30%	
			mAP_{50}	mAP_{75}	mAP_{50}	mAP_{75}	mAP_{50}	mAP_{75}
Supervised	Faster R-CNN* [55]	TPAMI 2016	50.79	-	53.35	-	55.69	-
	FCOS† [39]	ICCV 2019	56.07	-	59.98	-	62.60	-
Semi-supervised	Unbiased Teacher* [20]	ICLR 2021	57.53	21.03	59.11	22.80	61.01	24.27
	Soft Teacher* [21]	ICCV 2021	58.46	21.04	59.76	21.52	61.00	24.15
	Dense Teacher† [22]	ECCV 2022	59.32	16.60	62.28	21.25	64.45	25.03
	SOOD† [49]	CVPR 2023	59.55	18.09	62.43	21.02	64.52	24.30
	Adjacent Teacher†(**ours**)	-	**62.50**	**21.05**	**64.02**	**22.97**	**65.16**	**25.06**

TABLE V
QUANTITATIVE RESULTS ON FAIR1M DATASET UNDER THE SETTING OF 10% PARTIALLY LABELED DATA.

Setting	Method	mAP_{50}
Supervised	Faster R-CNN* [55]	65.24
	FCOS† [39]	66.09
Semi-supervised	Unbiased Teacher* [20]	69.68
	Soft Teacher* [21]	70.94
	Dense Teacher† [21]	70.71
	SOOD† [49]	70.77
	Adjacent Teacher†(**ours**) [49]	**71.42**

TABLE VI
QUANTITATIVE RESULTS ON DOTA-V1.5 DATASET UNDER THE CONFIGURATION OF FULLY LABELED DATA.

Method	Publication	mAP_{50}
Unbiased Teacher* [20]	ICLR 2021	63.82 $\xrightarrow{-1.09}$ 62.73
Soft Teacher* [21]	ICCV 2021	63.82 $\xrightarrow{+1.72}$ 65.54
Dense Teacher† [22]	ECCV 2022	66.23 $\xrightarrow{+0.55}$ 66.78
Adjacent Teacher†(**ours**)	-	66.23 $\xrightarrow{+1.16}$ **67.39**

implementations of rotated-Faster-RCNN and rotated-FCOS, respectively.

1) Comparison on DOTA-v1.5: For the experiments conducted on the DOTA-v1.5 dataset under the partially labeled data setting, quantitative results in terms of mAP_{50} and mAP_{75} metrics are presented in Table III. The proposed approach demonstrates optimal performance in all proportion configurations compared to its competitors. Specifically, at annotation proportions of 10%, 20%, and 30%, the proposed method achieves mAP_{50} scores of 51.26, 57.42, and 60.26, surpassing the supervised baseline by 6.63, 5.01, and 5.49 mAP_{50}, and outperforming the SOTA method by 1.31, 0.37, and 0.73 mAP_{50}, respectively. In terms of the more stringent mAP_{75} metric, our Adjacent Teacher also achieves the highest accuracy across all annotation proportions, surpassing the SOTA method by 0.61, 0.73, and 0.67 mAP_{75}, respectively. The experiments demonstrate that the proposed method effectively utilizes unannotated data, even in cases of extremely sparse annotations. Simultaneously, it achieves more precise oriented object localization.

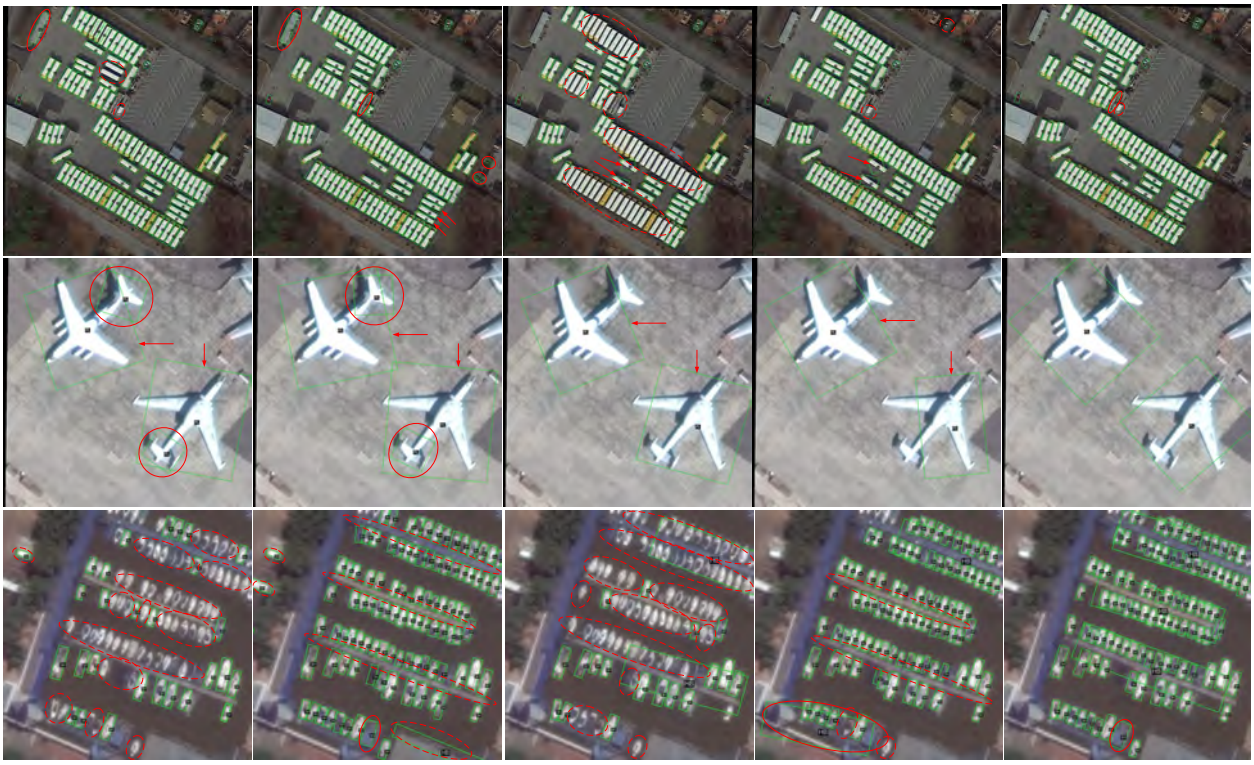
For the fully labeled data experiments, results of the compared methods and their respective baselines are presented in Table VI. Adjacent Teacher outperforms other approaches by a minimum of 0.61 mAP_{50}. In comparison to the baseline, an improvement of 1.16 points on the mAP_{50} metric is achieved.

Qualitative results of our Adjacent Teacher, compared to other SOTA models, are illustrated in Fig. 8. The detection results indicate that previous methods suffered from significant missed detections due to reliance exclusively on pseudo-labels with high confidence levels. Furthermore, numerous bounding boxes exhibited inaccurate orientations. With the assistance of the LPM and PAC modules, our approach effectively overcame the aforementioned issues, resulting in superior detection performance.

2) Comparison on SODA-A and FAIR1M: To assess the efficacy of our Adjacent Teacher across various datasets, a comparative evaluation is conducted on the SODA-A and FAIR1M datasets without hyper-parameters tuning.

Table IV suggests that our model attained the optimal performance across various evaluation metrics of the SODA-A dataset. Specifically, mAP_{50} scores of 62.50, 64.02, and 65.16 were achieved under different partially labeled data settings, respectively. In comparison to the baseline, our method improved by 6.43, 2.07, and 4.04 mAP_{50}, and outperformed the SOTA method by 2.95, 1.59, and 0.64 mAP_{50}, respectively. Moreover, it also achieved the best performance on the mAP_{75} metric, indicating that the bounding boxes generated by our method maintain higher precision and accuracy.

On the FAIR1M dataset under the 10% partially labeled setting (Table V), our Adjacent Teacher also achieves a SOTA



(a) Unbiased Teacher (b) Soft Teacher (c) Dense Teacher (d) SOOD (e) Ours

Fig. 8. Visualization of the detection results of different detection models from DOTA-v1.5 dataset with 10% partially labeled data setting. The detection results are represented by green boxes, false negative prediction are marked with red dashed region, false positive prediction are marked with red solid region, and inaccurate orientation prediction are marked with red arrow.

71.42 mAP_{50} , surpassing all compared semi-supervised methods. Specifically, it outperforms the previous best approach (Soft Teacher [21]) by 0.48 mAP_{50} and demonstrates 0.65 mAP_{50} improvement over the strong baseline SOOD [49].

Therefore, in the task related to oriented object detection within RSIs, the proposed method demonstrates excellent robustness and effectiveness.

3) Generalization on Other SSOD Models: To validate the universality of the proposed method, we apply our method to other existing SSOD models. As shown in Table VII, our method yields notable improvements across various models. Specifically, for the Unbiased Teacher [23], the proposed method elevates the mAP_{50} from 46.39 to 48.39, representing a significant enhancement of 2 points. In the case of the Soft Teacher [21], our method raises the mAP_{50} from 49.95 to 50.34, resulting in a gain of 0.39 points. Similarly, for the SOOD [49] method, our approach boosts the mAP_{50} from 49.67 to 51.26, indicating a notable enhancement of 1.59 points.

TABLE VII
THE EFFECTIVENESS OF THE PROPOSED METHOD ON OTHER SSOD METHODS UNDER THE SETTING OF 10% PARTIALLY LABELED DATA.

Method	Publication	mAP_{50}
Unbiased Teacher* [20]	ICLR 2021	46.39 $\xrightarrow{+2.00}$ 48.39
Soft Teacher* [21]	ICCV 2021	49.95 $\xrightarrow{+0.39}$ 50.34
SOOD† [49]	CVPR 2023	49.67 $\xrightarrow{+1.59}$ 51.26

TABLE VIII
ABLATION STUDIES OF EACH COMPONENT ON DOTA-V1.5 DATASET UNDER THE 10%, 20%, 30% PACTIALLY LABELED DATA SETTING.

Setting	Center-Sampling	LPM	PAC	mAP_{50}		
				10%	20%	30%
I	✓	-	-	49.67	55.90	58.86
II	-	-	-	50.11	56.29	59.09
III	-	✓	-	50.56	56.89	59.99
IV	-	-	✓	50.21	56.54	59.68
V	-	✓	✓	**51.26**	**57.42**	**60.26**

D. Ablation Studies

Here, we carry out extensive research on the DOTA-v1.5 dataset to verify the efficacy and robustness of our key design. Unless otherwise specified, all experiments are conducted on the setting of 10% partially labeled data setting.

1) Analysis of Each Component: We investigated the effects of the exclusion of Center-Sampling, the proposed Low-Confidence Pseudo-Label Mining (LPM), and Pseudo-Label Angle Correcting (PAC) modules on the experimental results. Note that the first row of Table VIII representing the SOOD [49] model excluding the Global Consistency loss module.

As depicted in Table VIII, excluding Center-Sampling from FCOS yields an improvement of at least 0.23 points in mAP_{50} across three distinct ratios of partially labeled data experimental configurations. Furthermore, the improvement becomes increasingly pronounced as the proportion of labeled

(a) Threshold:0.5 (b) Threshold:0.25 (c) With LPM (d) With PAR

Fig. 9. Visualization of the effect of LPM and PAC. The correct pseudo-labels are indicated by green bounding boxes, the missing pseudo-labels are marked within red dashed region, the false positive pseudo-labels are marked within red solid region, and inaccurate orientation pseudo-labels are marked with red arrow.

TABLE IX
THE IMPACT ON SPECIFIC CATEGORIES.

Setting	LPM	PAC	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	CC	mAP_{50}
I	-	-	**79.2**	**63.2**	27.6	**33.7**	44.5	65.1	77.3	87.0	40.0	54.5	**39.9**	58.6	45.3	**50.9**	33.9	0.0	50.11
II	✓	✓	78.9	63.1	**28.4**	33.3	**48.3**	**67.4**	**78.9**	**89.7**	**42.6**	**55.2**	39.5	**58.9**	**50.4**	50.6	**34.9**	0.0	**51.26**

TABLE X
THE EFFECT OF HYPER-PACAMETER λ FOR THE ADJACENT SPATIAL CONSTRAINTS.

Setting	λ	mAP_{50}
I	1.5	50.52
II	2.0	**51.26**
III	3.0	50.38
IV	4.0	50.19

data diminishes. These findings indicate that the issue of unreliable label assignments, arising from Center-Sampling in FCOS, simultaneously present in the tasks of semi-supervised horizontal and oriented object detection.

The third and fourth lines of Table VIII illustrate the effects of integrating the LPM and PAC modules into the model independently. The inclusion of the LPM leads to increases of 0.45, 0.6, and 0.9 points in mAP_{50} across various ratios of partially labeled data experiments. The addition of the PAC module results in improvements of 0.1, 0.25, and 0.59 points in mAP_{50} across different partially labeled data configurations. These results indicate that while the incorporation of a single module enhances performance, the effect remains relatively limited. This is because the LPM module aims to extract additional pseudo-labels that conform to adjacent spatial consistency constraints from low-confidence pseudo-labels, thereby increasing the quantity of generated pseudo-labels. Nevertheless, some of these low confidence pseudo labels lack accuracy, leading to limited performance gains. Conversely, the PAC module is designed to adjust the angles of low-confidence pseudo-labels to enhance their precision. Nonetheless, when applied independently, its impact remains somewhat constrained, as the pseudo-labels typically exhibit high confidence and accurate angles.

Ultimately, the simultaneous addition of all modules leads to

TABLE XI
THE EFFECT OF CATEGORY MAPPING FUNCTION M FOR THE CATEGORY CONSISTENCY CONSTRAINTS. {,} REPRESENTS GROUPING CATEGORIES WITHIN THE SET INTO A NEW CATEGORY.

Setting	M	mAP_{50}
I	Original Classification	50.61
II	{SV, LV}, {TC, BC}, {SH, HB}	**51.26**

TABLE XII
THE EFFECT OF HYPER-PACAMETER τ_{ang} FOR THE ANGLE CONSISTENCY CONSTRAINTS.

Setting	τ_{ang}	mAP_{50}
I	0.1	50.48
II	0.125	50.72
III	0.15	**51.26**
IV	0.175	50.48

TABLE XIII
THE AVERAGE TIME COST (IN SECONDS) OF EACH ITERATION DURING THE TRAINING PROCESS.

Setting	Method	Time Cost
I	Baseline	0.75
II	Baseline + LPM	0.77
III	Baseline + PAC	0.85
IV	Baseline + LPM + PAC	0.88

optimal model performance, with mAP_{50} increasing by 1.59, 1.52, and 1.4, respectively. The result further validates the module design and highlights the complementary interactions among them.

2) Analysis of the Impact on Specific Categories: To further analyze the impact of the proposed method on specific categories, we incorporated a detailed analysis of the detection performance for individual categories in the DOTA-v1.5 dataset within the ablation study. As shown in the results in Table IX, the detection performance significantly improved for categories such as small vehicles (SV), large vehicles (LV), ships (SH), tennis courts (TC), basketball courts (BC), and ports (HA), all of which are densely distributed and exhibit strong angle consistency. For categories such as baseball fields (BD), tracks (GTF), soccer fields (SBF), and roundabouts (RA), which have fewer instances and are more sparsely distributed, the detection performance remained unaffected, with no significant decrease in detection accuracy. This outcome aligns with the analysis presented in Table I, further validating the effectiveness of our method.

3) Visualizing the Effect of LPM and PAC: To provide an intuitive comprehension of the influence of the LPM and PAC modules, the complete process of pseudo-label mining and correction is presented in Fig. 9. It is obvious that applying a fixed confidence threshold for filtering pseudo-labels results in a large number of erroneous labels with a lower threshold, whereas a higher threshold leads to the omission of many accurate pseudo-labels. Moreover, both of them contain a significant number of bounding boxes with inaccurate angles. The LPM module effectively extracts pseudo-labels in alignment with the ASCP from low-confidence pseudo-labels, ensuring a lower error rate and minimizing the omission of accurate pseudo-labels. Similarly, the PAC module employs the ASCP to correct the orientations of pseudo-labels, thereby enhancing the precision of their bounding boxes.

4) Analysis of Hyper Parameter in Adjacent Spatial Constraints: In Adjacent Spatial Constraints, the λ determines the size of the adjacent spatial definition. Table X presents the results corresponding to varying values of λ. Setting λ to 2 yields the optimal performance. Performance declines with both larger and smaller values of λ. It is hypothesized that smaller values create an excessively limited local space, which restricts effective pseudo-label mining. Conversely, larger values create an overly broad local space, resulting in the introduction of more erroneous pseudo-labels.

5) Analysis of Category Mapping Function in Category Consistency Constraints: Closely related object categories should also be subject to category consistency constraints, not just limited to the same category. To verify it, the category mapping function M is utilized to respectively map large vehicle and small vehicle, ship and harbor, tennis court and basketball court to new categories. The results shown in Table XI suggest that remapping categories led to an improvement of 0.65 mAP_{50}, thereby validating it.

6) Analysis of Hyper Parameter in Angle Consistency Constraints: The parameter τ_{ang} controls the strictness of angle consistency. A higher value of τ_{ang} relaxes the angle constraint, allowing for the mining of more pseudo-labels that conform to ASCP. Nevertheless, it may also introduce more noisy pseudo-labels. Table XII illustrates that increasing τ_{ang} from 0.1 to 0.15 improves the performance of the method. However, further increasing it to 0.175 results in a slight performance decline. It is suggested that setting τ_{ang} to 0.15 achieves a favorable balance between the amount and accuracy of pseudo-labels, leading to optimal performance.

7) Analysis the Time Cost of LPM and PAC: The time cost of the proposed modules during training is detailed in Table XIII. Our proposed method introduces a process of mining and correcting pseudo-labels, leading to an increase in training time. The LPM module employs accelerated matrix operations to achieve low computational complexity, contributing negligible overhead to the overall training time. The PAR module requires iterative traversal of instances, resulting in higher computational complexity and more noticeable time consumption. In general, the increase in training time is relatively modest, at just 17%. And during the inference stage, there is no additional time cost, as processing pseudo-labels is not involved.

VI. Conclusion

Inspired by the First Law of Geography, this paper makes the first systematic identification and verification of the ubiquitous Adjacent Spatial Consistency Prior (ASCP) in remote sensing images. Building upon this fundamental discovery, we propose Adjacent Teacher, a novel semi-supervised oriented object detection framework with two innovative components: the Low-confidence Pseudo-label Mining (LPM) module and the Pseudo-label Angle Correction (PAC) module. LPM increases the number of reliable pseudo-labels by identifying low-confidence pseudo-labels that align with the ASCP, while

PAC improves the precision of pseudo-labels by adjusting their angles to meet the ASCP. Experimental results across multiple public datasets demonstrate that our method surpasses previous state-of-the-art approaches, exhibiting strong effectiveness and superiority. Furthermore, our method demonstrates strong applicability and can be seamlessly integrated into various semi-supervised oriented object detection frameworks, significantly enhancing their detection accuracy. Future work will focus on developing adaptive mechanisms to reduce the number of hyper-parameters, with particular focus on utilizing graph neural networks to model adjacent spatial information.

REFERENCES

[1] Z. Liu, J. Hu, L. Weng, and Y. Yang, "Rotated region based cnn for ship detection," in *Proceedings of the IEEE International Conference on Image Processing*, 2017, pp. 900–904.

[2] Y. Li, Q. Hou, Z. Zheng, M.-M. Cheng, J. Yang, and X. Li, "Large selective kernel network for remote sensing object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2023, pp. 16 794–16 805.

[3] J. Han, J. Ding, N. Xue, and G.-S. Xia, "Redet: A rotation-equivariant detector for aerial object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2786–2795.

[4] L. Gao, D. Wang, L. Zhuang, X. Sun, M. Huang, and A. Plaza, "Bs3lnet: A new blind-spot self-supervised learning network for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–18, 2023.

[5] L. Gao, X. Sun, X. Sun, L. Zhuang, Q. Du, and B. Zhang, "Hyperspectral anomaly detection based on chessboard topology," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.

[6] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. Plaza, "Bocknet: Blind-block reconstruction network with a guard window for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.

[7] J. Wang, M. Zhang, W. Li, and R. Tao, "A multistage information complementary fusion network based on flexible-mixup for hsi-x image classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 12, pp. 17 189–17 201, 2024.

[8] B. Zhao, M. Zhang, W. Li, X. Song, Y. Gao, Y. Zhang, and J. Wang, "Intermediate domain prototype contrastive adaptation for spartina alterniflora segmentation using multitemporal remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

[9] L. Hu, S. Li, J. Ruan, and F. Gao, "Semipsenet: A novel semi-supervised change detection network for remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

[10] X. Yang and J. Yan, "On the arbitrary-oriented object detection: Classification based approaches revisited," *International Journal of Computer Vision*, vol. 130, no. 5, pp. 1340–1365, 2022.

[11] X. Yang, L. Hou, Y. Zhou, W. Wang, and J. Yan, "Dense label encoding for boundary discontinuity free rotation detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15 819–15 829.

[12] X. Yang, J. Yan, Q. Ming, W. Wang, X. Zhang, and Q. Tian, "Rethinking rotated object detection with gaussian wasserstein distance loss," in *Proceedings of the International Conference on Machine Learning*, 2021, pp. 11 830–11 841.

[13] X. Yang, X. Yang, J. Yang, Q. Ming, W. Wang, Q. Tian, and J. Yan, "Learning high-precision bounding box for rotated object detection via kullback-leibler divergence," in *Proceedings of the Advances in Neural Information Processing Systems*, 2021, pp. 18 381–18 394.

[14] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, "Learning roi transformer for oriented object detection in aerial images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2849–2858.

[15] X. Yang, G. Zhang, W. Li, Y. Zhou, X. Wang, and J. Yan, "H2rbox: Horizontal box annotation is all you need for oriented object detection," in *Proceedings of the International Conference on Learning Representations*, 2023.

[16] Y. Yu, X. Yang, Q. Li, Y. Zhou, F. Da, and J. Yan, "H2rbox-v2: Incorporating symmetry for boosting horizontal box supervised oriented object detection," in *Proceedings of the Advances in Neural Information Processing Systems*, 2023, pp. 59 137–59 150.

[17] Y. Yu, X. Yang, Q. Li, F. Da, J. Dai, Y. Qiao, and J. Yan, "Point2rbox: Combine knowledge from synthetic visual patterns for end-to-end oriented object detection with single point supervision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2024, pp. 16 783–16 793.

[18] J. Luo, X. Yang, Y. Yu, Q. Li, J. Yan, and Y. Li, "Pointobb: Learning oriented object detection via single point supervision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2024, pp. 16 730–16 740.

[19] S. Zhang, J. Long, Y. Xu, and S. Mei, "Pmho: Point-supervised oriented object detection based on segmentation-driven proposal generation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–18, 2024.

[20] Y.-C. Liu, C.-Y. Ma, Z. He, C.-W. Kuo, K. Chen, P. Zhang, B. Wu, Z. Kira, and P. Vajda, "Unbiased teacher for semi-supervised object detection," in *Proceedings of the International Conference on Learning Representations*, 2021.

[21] M. Xu, Z. Zhang, H. Hu, J. Wang, L. Wang, F. Wei, X. Bai, and Z. Liu, "End-to-end semi-supervised object detection with soft teacher," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 3060–3069.

[22] H. Zhou, Z. Ge, S. Liu, W. Mao, Z. Li, H. Yu, and J. Sun, "Dense teacher: Dense pseudo-labels for semi-supervised object detection," in *Proceedings of the European Conference on Computer Vision*, 2022, pp. 35–50.

[23] Y.-C. Liu, C.-Y. Ma, and Z. Kira, "Unbiased teacher v2: Semi-supervised object detection for anchor-free and anchor-based detectors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9819–9828.

[24] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Economic Geography*, vol. 46, pp. 234–240, 1970.

[25] D.-H. Lee *et al.*, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proceedings of the International Conference on Machine Learning Workshops*, 2013, p. 896.

[26] Y. Grandvalet and Y. Bengio, "Semi-supervised learning by entropy minimization," in *Proceedings of the Advances in Neural Information Processing Systems*, 2004, pp. 529–536.

[27] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, "Self-training with noisy student improves imagenet classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 687–10 698.

[28] P. Bachman, O. Alsharif, and D. Precup, "Learning with pseudo-ensembles," in *Proceedings of the Advances in Neural Information Processing Systems*, 2014, pp. 3365–3373.

[29] M. Sajjadi, M. Javanmardi, and T. Tasdizen, "Regularization with stochastic transformations and perturbations for deep semi-supervised learning," in *Proceedings of the Advances in Neural Information Processing Systems*, 2016, pp. 1163–1171.

[30] T. Miyato, S.-i. Maeda, M. Koyama, and S. Ishii, "Virtual adversarial training: A regularization method for supervised and semi-supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1979–1993, 2018.

[31] K. Wang, X. Yan, D. Zhang, L. Zhang, and L. Lin, "Towards human-machine cooperation: Self-supervised sample mining for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1605–1613.

[32] J. Jeong, S. Lee, J. Kim, and N. Kwak, "Consistency-based semi-supervised learning for object detection," in *Proceedings of the Advances in Neural Information Processing Systems*, 2019, pp. 10 759–10 768.

[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," *arXiv preprint arXiv:2005.04757*, 2020.

[34] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proceedings of the Advances in Neural Information Processing Systems*, 2017, pp. 1195–1204.

[35] H. Li, Z. Wu, A. Shrivastava, and L. S. Davis, "Rethinking pseudo labels for semi-supervised object detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, 2022, pp. 1314–1322.

[36] Y. Tang, W. Chen, Y. Luo, and Y. Zhang, "Humble teachers teach better students for semi-supervised object detection," in *Proceedings of the*

IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 3132–3141.
[37] X. Wang, X. Yang, S. Zhang, Y. Li, L. Feng, S. Fang, C. Lyu, K. Chen, and W. Zhang, "Consistent-teacher: Towards reducing inconsistent pseudo-targets in semi-supervised object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3240–3249.
[38] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.
[39] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: Fully convolutional one-stage object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 9627–9636.
[40] X. Yang, J. Yan, Z. Feng, and T. He, "R3det: Refined single-stage detector with feature refinement for rotating object," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 4, 2021, pp. 3163–3171.
[41] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng, and X. Xue, "Arbitrary-oriented scene text detection via rotation proposals," *IEEE Transactions on Multimedia*, vol. 20, no. 11, pp. 3111–3122, 2018.
[42] M. Liao, Z. Zhu, B. Shi, G.-S. Xia, and X. Bai, "Rotation-sensitive regression for oriented scene text detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5909–5918.
[43] E. Xie, Y. Zang, S. Shao, G. Yu, C. Yao, and G. Li, "Scene text detection with supervised pyramid context network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 9038–9045.
[44] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, "Oriented r-cnn for object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 3520–3529.
[45] A. Khoreva, R. Benenson, J. Hosang, M. Hein, and B. Schiele, "Simple does it: Weakly supervised instance and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 876–885.
[46] C.-C. Hsu, K.-J. Hsu, C.-C. Tsai, Y.-Y. Lin, and Y.-Y. Chuang, "Weakly supervised instance segmentation using the bounding box tightness prior," in *Proceedings of the Advances in Neural Information Processing Systems*, 2019, pp. 6582–6593.
[47] Z. Tian, C. Shen, X. Wang, and H. Chen, "Boxinst: High-performance instance segmentation with box annotations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 5443–5452.
[48] Z. Tan, Z. Jiang, C. Guo, and H. Zhang, "Wsodet: A weakly supervised oriented detector for aerial object detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.
[49] W. Hua, D. Liang, J. Li, X. Liu, Z. Zou, X. Ye, and X. Bai, "Sood: Towards semi-supervised oriented object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2023, pp. 15 558–15 567.
[50] P. A. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, pp. 17–23, 1950.
[51] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "Dota: A large-scale dataset for object detection in aerial images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3974–3983.
[52] G. Cheng, X. Yuan, X. Yao, K. Yan, Q. Zeng, X. Xie, and J. Han, "Towards large-scale small object detection: Survey and benchmarks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 11, pp. 13 467–13 488, 2023.
[53] X. Sun, P. Wang, Z. Yan, F. Xu, R. Wang, W. Diao, J. Chen, J. Li, Y. Feng, T. Xu, M. Weinmann, S. Hinz, C. Wang, and K. Fu, "Fair1m: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 184, pp. 116–130, 2022.
[54] Y. Zhou, X. Yang, G. Zhang, J. Wang, Y. Liu, L. Hou, X. Jiang, X. Liu, J. Yan, C. Lyu, W. Zhang, and K. Chen, "Mmrotate: A rotated object detection benchmark using pytorch," in *Proceedings of the ACM International Conference on Multimedia*, 2022.
[55] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

Tao Xia is currently pursuing the Ph.D. degree in the School of Computer Science and the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include remote sensing image processing and deep learning.

Wei Jing received the B.M. degree in e-commerce and the M.S. degree in computer software and theory from Shandong University of Science and Technology, Qingdao, China, in 2019 and 2022 respectively. He is currently working toward the Ph.D. degree in the National Elite Institute of Engineering and the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include remote sensing image processing and deep learning.

Qi Wang received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision, machine learning, pattern recognition, and remote sensing. For more information, visit the link (https://crabwq.github.io/).