

# S<sup>2</sup>Teacher: Step-by-step Teacher for Sparsely Annotated Oriented Object Detection

Yu Lin<sup>1</sup>, Jianghang Lin<sup>1</sup>, Kai Ye<sup>1</sup>, You Shen<sup>1</sup>, Shengchuan Zhang<sup>1</sup>, Liujuan Cao<sup>1\*</sup>

<sup>1</sup>Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Ministry of Education of China, Xiamen University, 361005, P.R. China.

## Abstract

Although fully-supervised oriented object detection has made significant progress in remote sensing image understanding, it comes at the cost of labor-intensive annotation. Recent studies have explored weakly and semi-supervised learning to alleviate this burden. However, these methods overlook the difficulties posed by dense annotations in complex remote sensing scenes. In this paper, we introduce a novel setting called sparsely annotated oriented object detection (SAOOD), which only labels partial instances, and propose a solution to address its challenges. Specifically, we focus on two key issues in the setting: (1) **sparse labeling leading to overfitting on limited foreground representations**, and (2) **unlabeled objects (false negatives) confusing feature learning**. To this end, we propose the S<sup>2</sup>Teacher, a novel angle-consistency guided method that progressively mines pseudo-labels for unlabeled objects from easy to hard, enhancing foreground representations. Additionally, it reweights the loss of unlabeled objects to mitigate their impact during training. Extensive experiments demonstrate that S<sup>2</sup>Teacher not only significantly improves detector performance across different sparse annotation levels but also achieves near-fully-supervised performance on the DOTA dataset with only 10% annotation instances, effectively balancing accuracy and labeling cost.

**Code** — <https://github.com/YL-XMU/S2Teacher>

## Introduction

Oriented object detection has achieved great success in understanding the remote sensing images in recent years (Yang et al. 2023b; Yu et al. 2024a). However, its development is hindered by the high cost of annotation, as labeling a rotated box (RBox) is approximately 36.5% more expensive than labeling a horizontal box (HBox) (Yang et al. 2023a). To address this, recent studies have explored training oriented detectors using HBox supervision (Yang et al. 2023a; Yu et al. 2024c), point supervision (Luo et al. 2024), and semi-supervision (Hua et al. 2023) to reduce annotation costs. These methods have yielded promising results in oriented object detection. However, they overlook a key characteristic of remote sensing images: **the prevalence of densely**

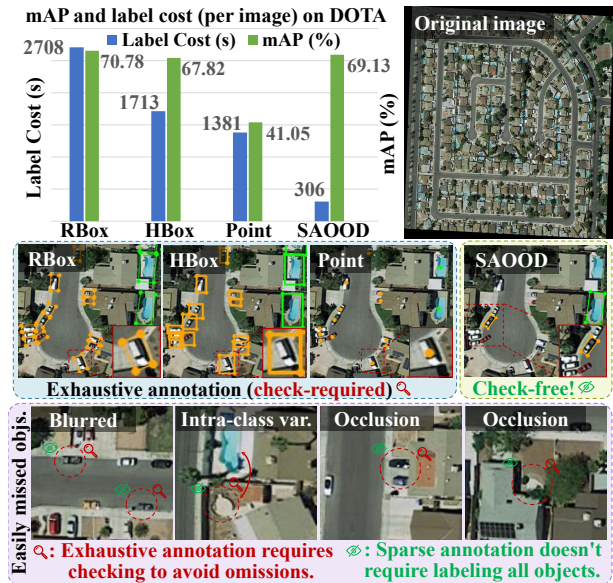


Figure 1: Compare annotation methods. RBox, HBox, and point supervision require labeling all objects and checking to avoid missed. In remote sensing, dense small objects and blurring/occlusion make this hard. Sparse annotation labels partial objects randomly, greatly reducing cost.

**labeled scenes.** As shown in Figure 1, the small size, partial occlusion, and blurred features of objects in remote sensing images make it **extremely challenging to label all objects without omission**. This requires annotators to repeatedly check to prevent missing objects, which is time-consuming. In fact, due to the difficulties associated with annotating dense, small objects (such as densely parked cars), many cars in DOTA-v1.0 (Ding et al. 2021) dataset are unlabeled.

As the saying goes, “Birds of a feather flock together.” Under a macroscopic remote sensing view, objects that exhibit spatial clustering are typically of the same class with similar features. Partial annotation of such objects can capture most of their features. As shown in Figure 1, we conducted experiments on the DOTA dataset. For a dense scene image containing 411 instances, annotating rotated

\*Corresponding author.

boxes (RBox) requires about 2708s, annotating horizontal boxes (HBox) takes about 1713s, and point annotation takes around 1381s. In contrast, sparse annotation of only 10% of the instances requires just 306s. Although HBox and point annotations reduce labeling time by 36.7% and 49% respectively, their efficiency gains are limited due to the time-consuming process of checking missed objects in dense scenes. In contrast, sparse annotation achieves a higher efficiency gain (88.7%) by not only reducing the number of labeled instances but also eliminating the need for exhaustive checking. Based on this observation, we naturally introduced a novel setting: sparsely annotated oriented object detection (SAOOD), which annotated partial instances. **The key advantage of SAOOD is:** 1) Unlike weakly and semi-supervised methods (Lin et al. 2024; Chen, Han, and DeBattista 2024), SAOOD enables random instance annotation in dense scenes, eliminating the need for repeated checks to avoid missed objects. 2) It also avoids the misleading supervision caused by label-omitted objects due to blurred features and occlusion in remote sensing images. 3) By distributing annotation across more images, SAOOD enhances feature diversity, whereas semi-supervised methods focus on a limited subset, leading to feature redundancy.

Although pseudo-label mining performs well in semi-supervised, applying it directly to SAOOD faces two major issues: 1) Unlabeled objects are misclassified as negatives, confusing the detector. 2) Complex dense scenes and diverse object orientations in remote sensing increase the difficulty of generating reliable pseudo-labels. Previous methods generate pseudo-labels for all instances at once, which in SAOOD adds significant noise. To address these, we propose a progressive pseudo-label generation framework, called S<sup>2</sup>Teacher. Specifically, S<sup>2</sup>Teacher includes three modules: Angle-Consistent Pseudo-label Generation (ACP), Entropy-Gaussian-based Pseudo-label Filtering (EGPF), and Pseudo-label Freezing (PLF). ACP generates reliable pseudo-labels through intra-cluster group decision and removes orientation error pseudo-labels via an angle-consistency mechanism. EGPF filters out false positive (FP) via Gaussian modeling of entropy, suppressing pseudo-label noise. PLF progressively freezes high-confidence pseudo-labels using temporal consistency, guiding S<sup>2</sup>Teacher to explore harder objects. Additionally, Focal Ignore Loss reweights supervision to mitigate confusion. Our main contributions are as follows:

- We analyze the key factor for the high annotation cost in remote sensing, namely that densely annotated scenes make exhaustive labeling difficult, and suggest introducing the SAOOD setting to address this challenge.
- We propose a novel framework, S<sup>2</sup>Teacher. ACP ensures reliable, angle-accurate pseudo-labels in dense scenes via clustering and angle consistency. EGPF suppresses pseudo-label noise through Gaussian modeling of entropy. PLF encourages the teacher to mine harder objects by freezing pseudo-labels. Focal Ignore Loss alleviates confusion caused by unlabeled objects.
- Experiments show that S<sup>2</sup>Teacher surpasses state-of-the-art methods and achieves near fully-supervised performance on DOTA using only 10% annotations.

## Related Work

**Oriented object detection.** Oriented object detection is widely used in remote sensing (Xu et al. 2024; Pu et al. 2023; Liu et al. 2025), with representative methods including two-stage Oriented R-CNN (Xie et al. 2021), one-stage S<sup>2</sup>A-Net (Han et al. 2021), and anchor-free Rotated FCOS (Tian et al. 2019). Prior work focused on model-level improvements such as feature alignment (Yang et al. 2021), angle handling (Xiao et al. 2024), and large-kernel convolution (Li et al. 2023). As large models emerge, data scale becomes crucial (Yao et al. 2025a,b; Liu et al. 2024). Yet in remote sensing, dense objects make full annotation costly.

**Low-cost oriented object detection.** Recent works reduce annotation costs in oriented object detection through three approaches: 1) Weakly supervised methods use HBox (Yang et al. 2023a) or point annotations (Luo et al. 2024; Yu et al. 2024b; Ren et al. 2024); point-based methods are cheaper but less accurate, while HBox-based methods preserve accuracy with limited savings. 2) Semi-supervised methods (Hua et al. 2023; Fang et al. 2024; Ma et al. 2025) balance cost and performance but overlook annotation density in remote sensing. 3) Weakly and semi-supervised methods (Wu et al. 2024b) combine RBox and point labels but still ignore density, limiting cost reduction. To address this, we propose SAOOD. Unlike semi-supervision, which uses fully labeled or unlabeled images, sparse annotation allows partial instance labeling (Wang et al. 2021). Though rarely used in natural scenes due to low object density (e.g., 7 objects per image in COCO), it suits remote sensing, where dense scenes (e.g., 143 per image in DOTA-v1.5) make detailed annotation costly and SAOOD more practical.

**Sparsely annotated object detection (SAOD).** Recent SAOD methods (Chen et al. 2025) have mainly focused on natural scenes. Co-mining (Wang et al. 2021) uses a Siamese network, Co-student (Wu et al. 2024a) employs two students for pseudo-labeling, SparseDet (Suri et al. 2023) applies self-supervised learning, and Calibrated Teacher (Wang et al. 2023) enhances confidence calibration. However, in remote sensing, dense scenes and angle parameters hinder pseudo-labeling, causing low-quality labels. Our method addresses this through progressive mining and angle-consistent pseudo-labeling for accurate orientation.

## S<sup>2</sup>Teacher

Given training images  $\mathcal{X}$ ,  $\mathcal{Y}_s$  represents the sparse annotated ground truths (GT). The goal of SAOOD is to use  $\{\mathcal{X}, \mathcal{Y}_s\}$  to initially train a model, mine pseudo labels  $\mathcal{Y}_p$  from the unlabeled object set, and then continue training with  $\{\mathcal{X}, \mathcal{Y}_s \cup \mathcal{Y}_p\}$ , thereby iteratively improving performance. The overall structure of S<sup>2</sup>Teacher is shown in Figure 2. The input image is weakly enhanced for the teacher model and strongly enhanced for the student model. The teacher model generates coarse pseudo labels using our proposed cluster-based angle consistency pseudo-label generation module (ACP). These pseudo-labels then pass through the entropy Gaussian modeling-guided pseudo-label filtering module (EGPF), resulting in refined pseudo-labels that are used for supervised training of the student model. In each

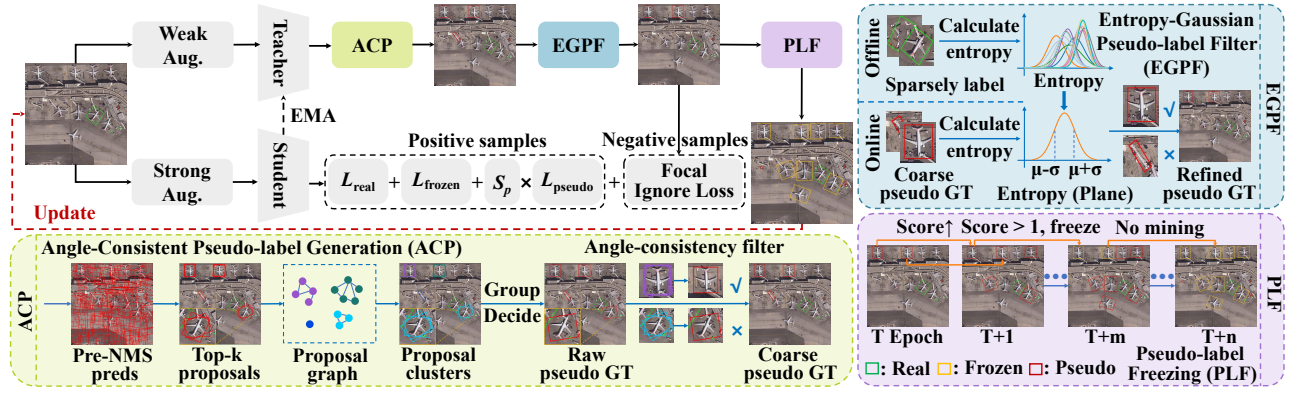


Figure 2: The overall framework of S<sup>2</sup>Teacher. The input image is processed by the teacher and ACP to prioritize easy-object mining. After false positives are filtered by EGPF, the pseudo GT supervises the student. Across iterations, PLF compares pseudo GTs, freezing high-confidence ones and encouraging ACP to mine harder unlabeled objects.

iteration, the refined pseudo-labels are sequentially compared using the pseudo-label freezing module (PLF), gradually freezing high-confidence pseudo-labels as real labels. This forces the teacher model to continue mining new, difficult pseudo-labels. The training loss of student models is divided into two parts: positive sample loss and negative sample loss. Positive sample loss includes real ground truth (GT), frozen GT, and pseudo GT loss. The negative sample loss adopts our designed Focal Ignore Loss.

### Angle-Consistent Pseudo-label Generation

In sparsely annotated remote sensing, pseudo-label generation faces two key challenges. 1) High object density and complex scenes make reliable pseudo-labels hard. Prior methods generate pseudo-labels for all objects per iteration using single-proposal predictions, producing numerous FP due to prediction randomness. 2) Oriented object detection introduces an angle parameter, expanding the solution space and increasing the risk of angular errors in pseudo labels, which can harm localization learning.

We propose angle-consistent pseudo-label generation (ACP). ACP uses stepwise mining, selecting the most confident (i.e., easiest) pseudo-labels at each iteration and gradually mining harder ones as the model improves. By generating pseudo-labels via group decisions within proposal clusters, ACP avoids randomness of individual proposal predictions and ensures reliability in dense scenes. To reduce angular errors, ACP applies an angle-consistency mechanism that improves angular accuracy via group decision and filters out pseudo-labels with incorrect angles based on angular consistency estimation, preventing harmful supervision.

As shown in Figure 2, the weakly augmented image is processed by the teacher model to produce pre-NMS outputs, which are then passed to ACP. ACP first filters out background proposals with low confidence (below 0.6) or significant overlap with sparsely annotated ground truths (GT), and then selects the Top- $k$  highest-confidence proposals. These proposals are used to construct a proposal graph, where each proposal is treated as a node. Inspired by (Tang

et al. 2018), two proposals with an IoU greater than 0.5 are considered connected, thus forming a proposal graph. Proposal clusters are then searched using a greedy algorithm, with each cluster representing a potential unlabeled object. The group decision within each cluster yields the raw pseudo-labels, reducing individual prediction randomness. Specifically, the pseudo-label confidence  $S_p$  is computed by jointly considering the number and classification scores of proposals within each cluster (more proposals suggest a higher likelihood of potential unlabeled objects). The pseudo-label category  $C_p$  is determined as the majority category among proposals in the cluster, as formulated below:

$$S_p = \max_{i \in \{1, N_p\}} (S_{c,i}) \cdot \frac{N_p}{k}, \quad C_p = \arg \max_{c \in \{1, 2, \dots, K\}} \sum_{i=1}^{N_p} \mathbb{1}_{\{c_i=c\}}, \quad (1)$$

where  $S_{c,i}$  is the classification score of the  $i$ -th proposal in the proposal cluster,  $N_p$  is the number of proposals in each cluster.  $K$  is the number of classes,  $\mathbb{1}$  is the indicator function, being 1 if  $c_i = c$  and 0 otherwise,  $c_i$  is the top-scoring category of the  $i$ -th proposal. Then, we calculate the weighted average position of the proposals within the cluster to obtain the pseudo bounding box  $B_p$ :

$$B_p(x, y, w, h) = \frac{1}{\sum_{i=1}^{N_p} S_{c,i}} \sum_{i=1}^{N_p} S_{c,i} \cdot B_{c,i}(x, y, w, h), \quad (2)$$

where  $B_{c,i}$  is the  $i$ -th proposal box in the cluster. Due to the angular discontinuity in oriented object detection (Yu and Da 2024), directly averaging angles within a cluster can cause numerical instability near boundary. To tackle this, all angles are first mapped from the 180° domain to 360° range and converted into unit vectors via  $\mathbf{v}_i = (\cos \theta_i, \sin \theta_i)$ . After computation, the angles are mapped back. The pseudo-label angle is defined as the angle of the average vector:

$$B_p(\theta) = \arctan 2 \left( \sum_{i=1}^{N_p} S_{c,i} \sin(\theta_i), \sum_{i=1}^{N_p} S_{c,i} \cos(\theta_i) \right), \quad (3)$$

where  $\theta_i$  is the angle of the  $i$ -th proposal. The norm of the average vector quantifies the angular consistency score  $S_\theta$ :

$$S_\theta = \left\| \frac{1}{\sum_{i=1}^{N_p} S_{c,i}} \sum_{i=1}^{N_p} S_{c,i} \mathbf{v}_i \right\|_2, \quad (4)$$

Higher  $S_\theta$  reflects consistent orientations in the proposal cluster, while large angular discrepancies reduce  $S_\theta$  due to vector cancellation. Raw pseudo labels with  $S_\theta < \sqrt{3}/2$  (angle difference  $> 30^\circ$  in the  $180^\circ$  domain) are discarded to avoid harming the localization learning of student, yielding coarse pseudo labels. Cluster-based group decision enhances label reliability in dense scenes, while angle-consistency-guided generation reduces angular errors and preserves localization accuracy.

### Pseudo-label Filtering

A major challenge in pseudo-label-based methods is the interference from false positive (FP) pseudo labels during training. Due to inevitable teacher prediction errors, FP pseudo labels can mislead the gradient direction. To address this, we designed an entropy Gaussian modeling-guided pseudo-label filtering module (EGPF), as shown in Figure 2. Information entropy is widely used to quantify the complexity and diversity of image regions (Mi et al. 2022). As defined in Equation 5, given a pixel probability distribution  $\{p_1, p_2, \dots, p_n\}$ , entropy reaches its maximum  $\log n$  when all probabilities are equal, and becomes zero when one probability  $p_i = 1$ . Background regions are generally smooth with limited pixel variation, leading to concentrated pixel distributions and low entropy. Foreground exhibit complex structures and diverse pixels, resulting in higher entropy. We define object information entropy  $\mathcal{H}$  as:

$$\mathcal{H} = - \sum_{i=1}^n p(x_i) \log(p(x_i)), \quad (5)$$

where  $n$  is the number of pixels in the box, and  $p(x_i)$  is the value distribution of each pixel in the box. Due to variations in material, texture, color, and other visual traits, object entropy differs across categories. Visually complex objects (e.g., ships) typically have higher entropy, while simpler ones (e.g., sports fields) show lower entropy. Therefore, EGPF computes the entropy of sparsely labeled GT offline and models a Gaussian distribution  $p(\mathcal{H}_c)$  for each category:

$$p(\mathcal{H}_c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{(\mathcal{H}_c - \mu_c)^2}{2\sigma_c^2}\right), \quad (6)$$

where  $\mu_c$  is the mean of  $\mathcal{H}_c$ , and  $\sigma_c^2$  is the variance of  $\mathcal{H}_c$ . This offline stage runs once before training and does not affect inference time. During the online training stage, EGPF calculates  $\mathcal{H}_{pseudo}$  of objects within the coarse pseudo GT mined by the ACP, and removes the FP pseudo labels by:

$$\text{Filter}(\mathcal{H}_{pseudo}) = \mathbb{1}_{\{\mu - \sigma \leq \mathcal{H}_{pseudo} \leq \mu + \sigma\}}, \quad (7)$$

where Filter is the filtering function,  $\mathbb{1}$  is the indicator function. When  $\mu - \sigma \leq \mathcal{H}_{pseudo} \leq \mu + \sigma$  is satisfied, the refined pseudo GT is retained, otherwise it is filtered out. This filter ensures the quality of refined pseudo-labels and prevents numerous FP pseudo-labels from misleading the student.

### Pseudo-label Freezing

To ensure high-quality pseudo-labels, ACP selects the Top- $k$  proposals with the highest scores for pseudo-label mining. However, this may lead to overlapping pseudo-labels in each iteration, preventing the teacher model from discovering new unlabeled objects. Therefore, we propose the pseudo-label freezing module (PLF), which complements the ACP to enable teacher model mine more difficult pseudo-labels.

The structure of PLF is shown in Figure 2. The refined pseudo-labels mined in each iteration are stored in a queue, with the queue length corresponding to the number of iterations per epoch. PLF calculates the IoU between the pseudo GT mined in the current epoch and those from the previous epoch to determine if they correspond to the same object. When a pseudo GT is mined multiple times at the same location, it indicates a high probability of an unlabeled object in that region. In this case, PLF increases the confidence of the pseudo GT. Conversely, when a pseudo GT is mined at a location in previous epochs but not in the current epoch, it suggests a decreased likelihood of an unlabeled object, prompting PLF to reduce the confidence of that pseudo GT. After each iteration, the mined pseudo-labels are stored in the queue following the first-in, first-out (FIFO) principle. If a location is repeatedly mined for pseudo GT, its confidence continues to increase. Once the confidence exceeds 1, it indicates a high probability of an unlabeled object at that location. At this point, PLF freezes the pseudo GT as a real GT, which is treated as real GT in the loss calculation and does not require mining. This process allows the ACP to focus on mining new unlabeled objects. In essence, PLF serves as a temporal group decision mechanism, aggregating pseudo-label mining results from multiple epochs to jointly assess the likelihood of an unlabeled object at a given location.

### The overall loss

**Total Loss.** The overall loss function of S<sup>2</sup>Teacher consists of positive and negative sample losses. The positive loss includes the loss from manually annotated real GT and the pseudo GT generated by the teacher model. The pseudo GT loss is divided into two components: the frozen pseudo GT, which is treated with the same weight (1.0) as real GT, and the ordinary pseudo GT, which is weighted by  $S_p$  from the ACP. The negative sample loss is using Focal Ignore Loss  $\mathcal{L}_{F-I}^{cls}$ . Therefore, the total loss can be formulated as:

$$\mathcal{L}_{total} = \mathcal{L}_{GT} + \mathcal{L}_{fz\text{GT}} + S_p \cdot \mathcal{L}_{pseu\text{GT}} + \mathcal{L}_{F-I}^{cls}, \quad (8)$$

where  $\mathcal{L}_{GT}$  is the loss of real GT,  $\mathcal{L}_{fz\text{GT}}$  is the loss of frozen GT,  $\mathcal{L}_{pseu\text{GT}}$  is the loss of pseudo GT, all of them are composed of classification and regression losses, Focal Loss (Lin et al. 2017) is used for classification, and IoU loss (Yu et al. 2016) is used for regression.

**Focal Ignore Loss.** In SAOOD, objects are partially labeled, causing proposals near unlabeled objects to be treated as negative samples despite sharing features with positives. This mislabeling confuses the detector, which mistakes foreground for background. One-stage, anchor-free detectors classify all proposals outside GT as negatives, generating more misleading negatives than positives, thus misdirecting gradients. Moreover, using Focal Loss (Lin et al. 2017)

treats these misleading samples as hard negatives, worsening the problem. To address this, we propose Focal Ignore Loss, which focuses on true hard negatives while ignoring misleading negatives. As shown in Equation 9, misleading negatives have features similar to positives, leading the teacher to assign them low background confidence; using this confidence as the loss weight prevents these samples from dominating training. However, hard negatives near real GT also have low background confidence, so simply down-weighting all low-confidence samples would hinder learning and cause foreground over-prediction. To avoid this, we compute the IoU between low-confidence proposals and GT, distinguishing hard negatives with high IoU from misleading negatives.

$$\mathcal{L}_{\text{F-I}}^{cls} = -\frac{1}{N_{hn}} \sum_{n=1}^{N_{hn}} \sum_{i=1}^C \alpha_i (1 - p_{n,i}^S)^\gamma \log(p_{n,i}^S) - \frac{1}{N_n} \cdot \sum_{m=1}^{N_n} (1 - q_m^T) \sum_{i=1}^C \alpha_i (1 - p_{m,i}^S)^\gamma \log(p_{m,i}^S) \quad (9)$$

where  $N_{hn}$  and  $N_n$  are the numbers of hard and normal negatives,  $C$  is the number of categories,  $\alpha_i$  and  $\gamma$  follow Focal Loss,  $p_{n,i}^S$  is the student-predicted probability for true class, and  $q_m^T$  is the teacher-predicted foreground probability.

## Experiments

### Datasets

We evaluate our method on three widely used remote sensing datasets: DOTA-v1.0 (Ding et al. 2021), DOTA-v1.5, and DIOR (Cheng et al. 2022). Since SAOOD requires sparsely annotated data, we construct sparse datasets by label sampling. Previous work (Lu et al. 2024) sampled a fixed proportion per class from the full dataset, assuming prior knowledge of the overall class distribution, which is unrealistic in practice. Instead, we adopt a more practical strategy (Wang et al. 2023; Suri et al. 2023) that randomly samples objects by category within each images. When the sampling number yields a non-integer, we round the result and retain at least one instance to avoid zero-shot cases. This approach better reflects real-world settings, where annotators can observe object counts per image and tend to label more frequent classes. Following standard practice (Yang et al. 2021), DOTA images are cropped into 1024×1024 patches. We report mean Average Precision (mAP) at an IoU threshold of 0.5. **DOTA-v1.0** is a remote sensing dataset with 2,806 images, 15 categories, 188,282 instances. **DOTA-v1.5** adds numerous tiny objects, averaging 143 objects per image, making it a dense scene dataset. **DIOR-R** consists of 23,463 images and 190,288 instances across 20 categories. It is challenging due to variations in object scale and weather.

### Experimental settings

Following the pseudo-label generation paradigm of SOOD (Hua et al. 2023), we adopt **Rotated FCOS as the baseline, trained using only sparse annotations**. The implementation and hyperparameters follow mmrotate (Zhou et al. 2022b). All models are trained on 4 RTX3090 GPUs using SGD with a learning rate of 0.0025, momentum of 0.9,

and weight decay of 0.0001. Following the teacher-student paradigm, the teacher is updated via EMA with a momentum of 0.9996. Since EMA typically improves performance, **all baselines are reported using EMA-updated teacher models to ensure fair comparison**, resulting in slightly higher mAP than conventional baselines without EMA.

### Main results

**SAOOD: less cost, better accuracy.** We compare the detection accuracy of state-of-the-art methods under different annotation settings on DOTA-v1.0. Note that semi-supervised methods annotate  $k\%$  of images, whereas SAOOD annotates  $k\%$  of instances per image. To unify annotation cost, we define the **Box Ratio** =  $N_A/N_{\text{total}}$ , where  $N_A$  is the number of annotated objects and  $N_{\text{total}}$  is the total number. For instance, 7.9% RBox in Table 1 means 7.9% objects are labeled. Since at least one instance is selected per image, **10% Annotation Ratio in Table 2 corresponds to 14% RBox in Table 1**.

As shown in Table 1, H2RBox (Yang et al. 2023a) achieves 67.82% mAP, though HBox annotations remain costly (Figure 1). Point supervision lowers cost but drops accuracy, with PointOBB-v2 (Ren et al. 2024) at only 44.85% mAP. Semi-supervised methods are sensitive to annotation amount: S<sup>2</sup>O-Det (Fu et al. 2024) achieves 55.18% mAP at 7.9% RBox and 67.70% at 27.6%. Under the same 7.9% RBox, S<sup>2</sup>Teacher achieves 64.59% mAP, and 69.13% mAP at 14%, surpassing S<sup>2</sup>O-Det with 27.6% RBox, showing superior performance with nearly half the annotation. We argue that focusing annotations on a subset of images is inefficient, as objects often share redundant features. SAOOD distributes annotations across more images, enabling broader feature learning. Despite using one-stage FCOS, S<sup>2</sup>Teacher outperforms PECL (Lu et al. 2024), built on the two-stage ReDet. This mainly stems from the cluster-based group decision, which improves pseudo-label reliability in dense scenes compared to the single-proposal decision used by PECL. The easy-to-hard mining strategy further enhances stability. Moreover, PECL ignores angular errors of pseudo-labels, S<sup>2</sup>Teacher employs an angle-consistency mechanism to improve angular accuracy. Other methods: Point2RBox (Yu et al. 2024b), PointOBB (Luo et al. 2024), Point2RBox-v2 (Yu et al. 2025), and SOOD++ (Liang et al. 2024).

**Gains on DOTA-v1.0.** As shown in Table 2, S<sup>2</sup>Teacher significantly outperforms the baseline under various annotation ratios. With 1% annotations, it surpasses the baseline by 7.42% mAP, attributed to its progressive pseudo-label mining that gradually uncovers unlabeled objects from easy to hard. At 10% annotations, S<sup>2</sup>Teacher achieves 69.13% mAP, approaching the fully supervised Rotated FCOS (70.78%). S<sup>2</sup>Teacher is particularly effective for densely distributed objects like large vehicles (LV). Under 1% annotations, the mAP for LV improves from 50.9% to 66.6% (+15.7%). This gain stems from angle-consistent pseudo-label generation, which improves orientation prediction for high aspect ratio objects (LV), and the cluster-based strategy, which enhances pseudo-label quality in dense scenes. S<sup>2</sup>Teacher also works well with two-stage detectors (e.g., Oriented R-CNN).

**Performance in dense scenarios on DOTA-v1.5.** DOTA-v1.5 is a dense dataset (avg. 143 objects per image), making

Annotation method	Method	Box Ratio	mAP(%)
RBox-supervised	Rotated FCOS $\diamond$	100% RBox	70.78
HBox-supervised	H2RBox $\diamond$	100% HBox	67.82
Point-supervised	Point2RBox $\ddagger$	100% Point	41.05
	PointOBB $\diamond$		30.08
	PointOBB-v2 $\clubsuit$		44.85
	Point2RBox-v2 $\diamond$		51.00
Semi-supervised	SOOD++ $\diamond$	7.9% RBox	54.17
	S <sup>2</sup> O-Det $\diamond$		55.18
	SOOD++ $\diamond$	27.6% RBox	64.93
Weakly and semi	(Wu et al. 2024b) $\diamond$	7.9%RBox+92.1%Point	59.69
	(Wu et al. 2024b) $\diamond$	27.6%RBox+72.4%Point	67.04
SAOOD	PECL $\clubsuit$	7.9% RBox	62.25
	S <sup>2</sup> Teacher $\diamond$ (Ours)		<b>64.59</b>
	PECL $\clubsuit$	14.0% RBox	68.12
	S <sup>2</sup> Teacher $\diamond$ (Ours)		<b>69.13</b>

Table 1: Comparison of state-of-the-art methods on DOTA-v1.0.  $\diamond$  means based on FCOS,  $\ddagger$  (YOLOF),  $\clubsuit$  (ReDet).

Annotation ratio	Method	mAP(%)
1%	Baseline (Rotated FCOS)*	57.17
	S <sup>2</sup> Teacher (Rotated FCOS-based)*	<b>64.59</b>
	Baseline (Oriented R-CNN) $\Delta$	55.82
	S <sup>2</sup> Teacher (Oriented R-CNN-based) $\Delta$	<b>66.21</b>
5%	Rotated FCOS*	60.11
	S <sup>2</sup> Teacher (Rotated FCOS-based)*	<b>67.62</b>
	Oriented R-CNN $\Delta$	61.07
	S <sup>2</sup> Teacher (Oriented R-CNN-based) $\Delta$	<b>68.98</b>
10%	Rotated FCOS*	63.30
	S <sup>2</sup> Teacher (Rotated FCOS-based)*	<b>69.13</b>
	Oriented R-CNN $\Delta$	64.29
	S <sup>2</sup> Teacher (Oriented R-CNN-based) $\Delta$	<b>70.06</b>

Table 2: The results of S<sup>2</sup>Teacher on DOTA-v1.0.  $\star$  means based on one-stage detector (Rotated FCOS), and  $\Delta$  means based on two-stage detector (Oriented R-CNN).

pseudo-label generation difficult due to visual complexity. Existing methods fall into two categories based on pseudo-label sparsity: sparse pseudo labels (SPL) (Liu et al. 2021; Fang et al. 2024), which use post-processing (e.g., thresholding, NMS), and dense pseudo labels (DPL) (Zhou et al. 2022a; Liang et al. 2024), which directly use dense outputs without NMS. Both generate pseudo labels for all instances at once, which is suboptimal for dense scenes and often leads to many FP pseudo labels. As shown in Table 3, these methods bring limited gains under different annotation ratios. In contrast, S<sup>2</sup>Teacher improves performance across all annotation ratios. This is due to its progressive mining strategy, which starts with high-confidence, easy instances (e.g., sports fields in Figure 3(b)) and gradually discovers harder ones (e.g., cars), enhancing stability in dense scenes. The cluster-based group decision avoids prediction randomness. Figure 3(a) shows pseudo labels mined in dense scenes.

**Good generalization on DIOR-R.** Table 4 shows that S<sup>2</sup>Teacher greatly outperforms the baseline on DIOR. With 10% annotations, it achieves 60.3% mAP (+6.4%), at 20%, it reaches 62.4% mAP, which is 99% of the fully supervised

Annotation ratio	Method	mAP(%)
1%	Baseline (Rotated FCOS)	51.50
	SPL	53.22
	DPL	51.78
	S <sup>2</sup> Teacher(Ours)	<b>59.59</b>
5%	Baseline (Rotated FCOS)	56.75
	SPL	56.83
	DPL	57.01
	S <sup>2</sup> Teacher(Ours)	<b>63.13</b>
10%	Baseline (Rotated FCOS)	59.37
	SPL	59.59
	DPL	59.59
	S <sup>2</sup> Teacher(Ours)	<b>63.52</b>

Table 3: Compare different pseudo-label generation methods on dense dataset DOTA-v1.5 (avg. 143 objects per image).

Annotation ratio	Method	mAP(%)
10%	Baseline (Rotated FCOS)	53.9
	S <sup>2</sup> Teacher (Ours)	<b>60.3</b>
20%	Baseline (Rotated FCOS)	56.1
	S <sup>2</sup> Teacher (Ours)	<b>62.4</b>
100%	Baseline (Rotated FCOS)	63.2

Table 4: Performance evaluation on DIOR-R dataset.

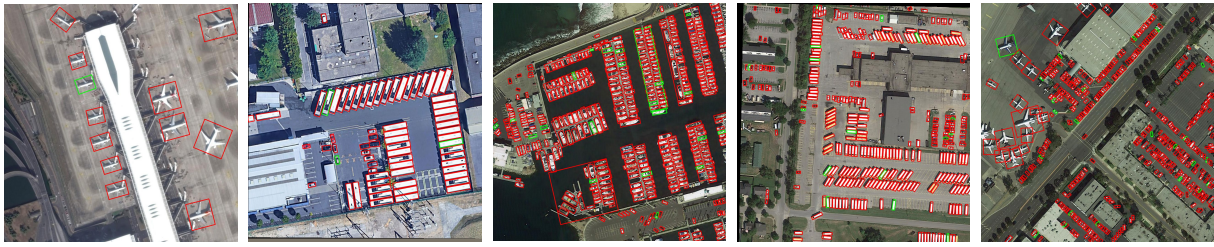
performance. These results highlight its strong generalization to aerial images under varying object scales.

**Better suited to OBB than HBB task.** We also evaluate S<sup>2</sup>Teacher on the horizontal bounding box (HBB) task of DOTA-v1.0. As shown in Table 5, it improves performance across annotation ratios, though the gains are more notable on the oriented bounding box (OBB) task. At 5% annotation, it yields improvements of 4.8% on HBB and 7.5% on OBB. This is attributed to the cluster-based pseudo-label generation guided by angle consistency, improving localization and angular accuracy. As OBB requires finer localization, the method brings greater benefit for OBB by preventing inaccurate pseudo labels from misleading model.

**Outperforming other sparse annotation methods in remote sensing.** Table 6 shows S<sup>2</sup>Teacher achieves the highest mAP, especially on one-stage detectors. In remote sensing, dense scenes and angle parameters make pseudo-labeling difficult, causing natural scene SAOD methods to produce inaccurate or poorly localized labels. S<sup>2</sup>Teacher uses progressive group-decision strategy to produce high-quality pseudo labels and angle-consistent generation to prevent angle errors from disrupting training. Two-stage detectors handle sparse annotations better by filtering unlabeled objects in RPN, while one-stage detectors treat all non-overlapping proposals as negatives, confusing learning. Our Focal Ignore Loss reduces this issue, narrowing their performance gap.

### Ablation and hyperparameter analysis

We conducted ablation studies on S<sup>2</sup>Teacher using 10% annotated DOTA-v1.0. As shown in Table 7, applying Focal Ignore Loss alone yields a limited mAP gain of 0.48%, as it mitigates the impact of misleading samples but fails to enhance foreground representation. Combining it with ACP brings a notable 4.76% improvement, as ACP progressively



(a) Pseudo labels mined during the training process.



(b) Visualization of PLF freeze pseudo-label process in different iterations.



(c) Manually missed objects.

Figure 3: Visualization of pseudo-label mining. Green box: manually annotated real GT; Red box: pseudo GT; Orange box: frozen pseudo GT by PLF; Blue box: object missed during manual annotation but mined by S<sup>2</sup>Teacher.

Annotation Ratio	Method	mAP(%)	
		HBB	OBB
5%	Baseline (FCOS)	57.1	60.1
	S <sup>2</sup> Teacher	61.9 (+4.8)	67.6 ( <b>+7.5</b> )
10%	Baseline (FCOS)	58.1	63.3
	S <sup>2</sup> Teacher	62.4 (+4.3)	69.1 ( <b>+5.8</b> )

Table 5: Comparison of OBB and HBB task performance.

Annotation method	Method	mAP(%)
Sparsely-annotated	Calibrated Teacher <sup>▲</sup> (Wang et al. 2023)	55.81
	Co-mining <sup>◦</sup> (Wang et al. 2021)	65.35
	SparseDet <sup>◦</sup> (Suri et al. 2023)	65.71
	PECL (Lu et al. 2024) (S <sup>2</sup> A-Net-based) <sup>▲</sup>	57.42
	PECL (ReDet-based) <sup>◦</sup>	67.06
	S <sup>2</sup> Teacher (Rotated FCOS-based) <sup>▲</sup>	67.62
	S <sup>2</sup> Teacher (Oriented R-CNN-based) <sup>◦</sup>	<b>68.98</b>

Table 6: Compare with other sparse annotation methods on the DOTA-v1.0 with 5% sparse annotation. ▲ means based on one-stage detector, ◦ means based on two-stage detector.

mines pseudo labels to enrich foreground features, while Focal Ignore Loss further suppresses misleading gradients, facilitating better mining. Adding EGPF increases the mAP to 68.87% by filtering out FP pseudo labels and improving label quality. Finally, incorporating PLF boosts the mAP to 69.13%, as it compares pseudo labels across epochs and freezes high-confidence ones. This temporal group decision ensures label quality and encourages ACP to explore harder objects, enabling step-by-step learning and sustained performance gains. We conducted experiments on the ACP hyperparameters. As shown in Table 5 of the supplementary material, varying the score threshold and Top-k values results in small mAP fluctuation (within 1%), indicating that ACP is not sensitive to hyperparameters. The best performance is

Focal Ignore Loss	ACP	EGPF	PLF	mAP(%)
				63.30
✓				63.78
✓	✓			68.06
✓	✓	✓		68.87
✓	✓	✓	✓	<b>69.13</b>

Table 7: The ablation study of each module in S<sup>2</sup>Teacher.

achieved when the score threshold is set to 0.6 and Top-k to 30. More detailed ablation and hyperparameter studies are provided in the supplementary material.

## Visualization Analysis

Figure 3(a) shows that S<sup>2</sup>Teacher accurately mines unlabeled objects (red boxes). Cluster-based progressive mining ensures reliable pseudo labels (e.g., ships and small vehicles) in dense scenes, while angle consistency enhances orientation accuracy for high aspect ratio objects (e.g., large vehicles). Figure 3(b) shows that PLF progressively freezes confident and simple objects (e.g., sports field), encouraging S<sup>2</sup>Teacher to further mine harder ones (e.g., small vehicles). Interestingly, as shown in Figure 3(c), S<sup>2</sup>Teacher discovers even manually missed objects (e.g., bridges), highlighting annotation difficulty and robustness of our method.

## Conclusion

We explore Sparsely Annotated Oriented Object Detection (SAOOD), a crucial yet underexplored task for reducing annotation costs in remote sensing. To tackle its challenges, we propose S<sup>2</sup>Teacher, which progressively mines pseudo labels via cluster-based group decision and angle consistency for accurate labeling in dense scenes. Focal Ignore Loss further mitigates misleading negatives. Experiments demonstrate the state-of-the-art performance of S<sup>2</sup>Teacher.

## Acknowledgments

This work was supported by the National Science Fund for Distinguished Young Scholars (No.62025603 and No.62525605), National Natural Science Foundation of China (No. U21B2037, U22B2051, No. U23A20383, No. 62176222, No. 62176226, No. 62272401, No. 62576300).

## References

- Chen, C.; Han, J.; and Debattista, K. 2024. Virtual category learning: A semi-supervised learning method for dense prediction with extremely limited labels. *IEEE transactions on pattern analysis and machine intelligence*, 46(8): 5595–5611.
- Chen, G.; Mao, Z.; Shen, J.; and Cheng, Z. 2025. Pseudo Label Guided Object Detection in Sparsely Annotated Underwater Optical Images. *IEEE Transactions on Geoscience and Remote Sensing*.
- Cheng, G.; Wang, J.; Li, K.; Xie, X.; Lang, C.; Yao, Y.; and Han, J. 2022. Anchor-free oriented proposal generator for object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–11.
- Ding, J.; Xue, N.; Xia, G.-S.; Bai, X.; Yang, W.; Yang, M. Y.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; et al. 2021. Object detection in aerial images: A large-scale benchmark and challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 44(11): 7778–7796.
- Fang, Z.; Ren, J.; Zheng, J.; Chen, R.; and Zhao, H. 2024. Dual teacher: Improving the reliability of pseudo labels for semi-supervised oriented object detection. *IEEE Transactions on Geoscience and Remote Sensing*.
- Fu, R.; Yan, S.; Chen, C.; Wang, X.; Heidari, A. A.; Li, J.; and Chen, H. 2024. S<sup>2</sup>O-Det: A Semisupervised Oriented Object Detection Network for Remote Sensing Images. *IEEE Transactions on Industrial Informatics*.
- Han, J.; Ding, J.; Li, J.; and Xia, G.-S. 2021. Align deep features for oriented object detection. *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, 60: 1–11.
- Hua, W.; Liang, D.; Li, J.; Liu, X.; Zou, Z.; Ye, X.; and Bai, X. 2023. Sood: Towards semi-supervised oriented object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 15558–15567.
- Li, Y.; Hou, Q.; Zheng, Z.; Cheng, M.-M.; Yang, J.; and Li, X. 2023. Large Selective Kernel Network for Remote Sensing Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 16794–16805.
- Liang, D.; Hua, W.; Shi, C.; Zou, Z.; Ye, X.; and Bai, X. 2024. SOOD++: Leveraging Unlabeled Data to Boost Oriented Object Detection. *arXiv preprint arXiv:2407.01016*.
- Lin, J.; Shen, Y.; Wang, B.; Lin, S.; Li, K.; and Cao, L. 2024. Weakly supervised open-vocabulary object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 3404–3412.
- Lin, T.-Y.; Goyal, P.; Girshick, R. B.; He, K.; Hariharan, B.; and S., D. D. M. 2017. Focal Loss for Dense Object Detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2999–3007.
- Liu, F.; Chen, D.; Guan, Z.; Zhou, X.; Zhu, J.; Ye, Q.; Fu, L.; and Zhou, J. 2024. Remoteclip: A vision language foundation model for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–16.
- Liu, F.; Yao, L.; Zhang, C.; Wu, T.; Zhang, X.; Jiang, X.; and Zhou, J. 2025. Boost UAV-based Object Detection via Scale-Invariant Feature Disentanglement and Adversarial Learning. *IEEE Transactions on Geoscience and Remote Sensing*.
- Liu, Y.-C.; Ma, C.-Y.; He, Z.; Kuo, C.-W.; Chen, K.; Zhang, P.; Wu, B.; Kira, Z.; and Vajda, P. 2021. Unbiased Teacher for Semi-Supervised Object Detection. *arXiv preprint arXiv:2102.09480*.
- Lu, Z.; Wang, C.; Xu, C.; Zheng, X.; and Cui, Z. 2024. Progressive Exploration-Conformal Learning for Sparsely Annotated Object Detection in Aerial Images. In *Proceedings of the Thirty-Eighth Annual Conference on Neural Information Processing Systems (NeurIPS)*.
- Luo, J.; Yang, X.; Yu, Y.; Li, Q.; Yan, J.; and Li, Y. 2024. PointOBB: Learning Oriented Object Detection via Single Point Supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16730–16740.
- Ma, Y.; Zhou, X.; Lan, S.; Wang, W.; Sun, Z.; and Qiao, Y. 2025. RemoteDPL: A Semi-supervised Object Detector with Dense Pseudo-Labels for Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing*.
- Mi, P.; Lin, J.; Zhou, Y.; Shen, Y.; Luo, G.; Sun, X.; Cao, L.; Fu, R.; Xu, Q.; and Ji, R. 2022. Active teacher for semi-supervised object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14482–14491.
- Pu, Y.; Wang, Y.; Xia, Z.; Han, Y.; Wang, Y.; Gan, W.; Wang, Z.; Song, S.; and Huang, G. 2023. Adaptive rotated convolution for rotated object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 6589–6600.
- Ren, B.; Yang, X.; Yu, Y.; Luo, J.; and Deng, Z. 2024. Pointobb-v2: Towards simpler, faster, and stronger single point supervised oriented object detection. *arXiv preprint arXiv:2410.08210*.
- Suri, S.; Rambhatla, S.; Chellappa, R.; and Shrivastava, A. 2023. Sparsedet: Improving sparsely annotated object detection with pseudo-positive mining. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6770–6781.
- Tang, P.; Wang, X.; Bai, S.; Shen, W.; Bai, X.; Liu, W.; and Yuille, A. 2018. PCL: Proposal Cluster Learning for Weakly Supervised Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 42(1): 176–191.
- Tian, Z.; Shen, C.; Chen, H.; and He, T. 2019. FCOS: Fully Convolutional One-Stage Object Detection. In *Proceedings*

- of the *IEEE/CVF International Conference on Computer Vision (ICCV)*, 6568–6577.
- Wang, H.; Liu, L.; Zhang, B.; Zhang, J.; Zhang, W.; Gan, Z.; Wang, Y.; Wang, C.; and Wang, H. 2023. Calibrated teacher for sparsely annotated object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 37, 2519–2527.
- Wang, T.; Yang, T.; Cao, J.; and Zhang, X. 2021. Co-mining: Self-supervised learning for sparsely annotated object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 35, 2800–2808.
- Wu, L.; Han, J.; Zheng, Z.; and Wang, X. 2024a. Co-Student: Collaborating Strong and Weak Students for Sparsely Annotated Object Detection. In *European Conference on Computer Vision*, 459–475. Springer.
- Wu, W.; Wong, H.-S.; Wu, S.; and Zhang, T. 2024b. Relational Matching for Weakly Semi-Supervised Oriented Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 27800–27810.
- Xiao, Z.; Yang, G.; Yang, X.; Mu, T.; Yan, J.; and Hu, S. 2024. Theoretically Achieving Continuous Representation of Oriented Bounding Boxes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16912–16922.
- Xie, X.; Cheng, G.; Wang, J.; Yao, X.; and Han, J. 2021. Oriented R-CNN for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 3520–3529.
- Xu, H.; Liu, X.; Xu, H.; Ma, Y.; Zhu, Z.; Yan, C.; and Dai, F. 2024. Rethinking boundary discontinuity problem for oriented object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17406–17415.
- Yang, X.; Yan, J.; Feng, Z.; and He, T. 2021. R3det: Refined single-stage detector with feature refinement for rotating object. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 35, 3163–3171.
- Yang, X.; Zhang, G.; Li, W.; Wang, X.; Zhou, Y.; and Yan, J. 2023a. H2rbox: Horizontal box annotation is all you need for oriented object detection. In *Proceedings of the International Conference on Learning Representations (ICLR)*.
- Yang, X.; Zhou, Y.; Zhang, G.; Yang, J.; Wang, W.; Yan, J.; Zhang, X.; and Tian, Q. 2023b. The KFIOU Loss for Rotated Object Detection. In *International Conference on Learning Representations (ICLR)*.
- Yao, L.; Liu, F.; Chen, D.; Zhang, C.; Wang, Y.; Chen, Z.; Xu, W.; Di, S.; and Zheng, Y. 2025a. Remotesam: Towards segment anything for earth observation. In *Proceedings of the 33rd ACM International Conference on Multimedia*, 3027–3036.
- Yao, L.; Liu, F.; Lu, H.; Zhang, C.; Min, R.; Xu, S.; Di, S.; and Peng, P. 2025b. RemoteReasoner: Towards Unifying Geospatial Reasoning Workflow. *arXiv preprint arXiv:2507.19280*.
- Yu, H.; Tian, Y.; Ye, Q.; and Liu, Y. 2024a. Spatial Transform Decoupling for Oriented Object Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, 6782–6790.
- Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; and Huang, T. 2016. Unitbox: An Advanced Object Detection Network. In *Proceedings of the 24th ACM International Conference on Multimedia (ACM MM)*, 516–520.
- Yu, Y.; and Da, F. 2024. On boundary discontinuity in angle regression based arbitrary oriented object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Yu, Y.; Ren, B.; Zhang, P.; Liu, M.; Luo, J.; Zhang, S.; Da, F.; Yan, J.; and Yang, X. 2025. Point2rbox-v2: Rethinking point-supervised oriented object detection with spatial layout among instances. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 19283–19293.
- Yu, Y.; Yang, X.; Li, Q.; Da, F.; Dai, J.; Qiao, Y.; and Yan, J. 2024b. Point2RBox: Combine Knowledge from Synthetic Visual Patterns for End-to-end Oriented Object Detection with Single Point Supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16783–16793.
- Yu, Y.; Yang, X.; Li, Q.; Zhou, Y.; Da, F.; and Yan, J. 2024c. H2RBox-v2: Incorporating symmetry for boosting horizontal box supervised oriented object detection. *Advances in Neural Information Processing Systems (NeurIPS)*, 36.
- Zhou, H.; Ge, Z.; Liu, S.; Mao, W.; Li, Z.; Yu, H.; and Sun, J. 2022a. Dense Teacher: Dense Pseudo-Labels for Semi-Supervised Object Detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 35–50. Springer.
- Zhou, Y.; Yang, X.; Zhang, G.; Wang, J.; Liu, Y.; Hou, L.; Jiang, X.; Liu, X.; Yan, J.; Lyu, C.; et al. 2022b. Mmrotate: A rotated object detection benchmark using pytorch. In *Proceedings of the 30th ACM International Conference on Multimedia*, 7331–7334.