

Partial Weakly-Supervised Oriented Object Detection

Mingxin Liu¹, Peiyuan Zhang², Yuan Liu¹, Wei Zhang¹, Yue Zhou³, Ning Liao¹
Ziyang Gong¹, Junwei Luo², Zhirui Wang⁴, Yi Yu^{5†}, Xue Yang^{1†}

¹SAIS & SCS, Shanghai Jiao Tong University ²Wuhan University ³East China Normal University
⁴Aerospace Information Research Institute ⁵Southeast University

<https://github.com/VisionXLab/PWOOD>

Abstract

The growing demand for oriented object detection (OOD) across various domains has driven significant research in this area. However, the high cost of dataset annotation remains a major concern. Current mainstream OOD algorithms can be mainly categorized into three types: (1) fully supervised methods using complete oriented bounding box (OBB) annotations, (2) semi-supervised methods using partial OBB annotations, and (3) weakly supervised methods using weak annotations such as horizontal boxes or points. However, these algorithms inevitably increase the cost of models in terms of annotation speed or annotation cost. To address this issue, we propose: (1) **the first Partial Weakly-Supervised Oriented Object Detection (PWOOD) framework** based on partially weak annotations (horizontal boxes or single points), which can efficiently leverage large amounts of unlabeled data, significantly outperforming weakly supervised algorithms trained with partially weak annotations, also offers a lower cost solution; (2) **Orientation-and-Scale-aware Student (OS-Student) model** capable of learning orientation and scale information with only a small amount of orientation-agnostic or scale-agnostic weak annotations; and (3) **Class-Agnostic Pseudo-Label Filtering strategy (CPF)** to reduce the model's sensitivity to static filtering thresholds. Comprehensive experiments on DOTA-v1.0/v1.5/v2.0 and DIOR datasets demonstrate that our PWOOD framework performs comparably to, or even surpasses traditional semi-supervised algorithms.

1. Introduction

In the field of oriented object detection, fully supervised learning [7, 40, 43, 44, 51] based on rotated bounding box

[†]Corresponding author. The work was in part supported by National Natural Science Foundation of China (62506229, 62306069), Natural Science Foundation of Shanghai under 25ZR1402268, and Shanghai QiYuan Innovation Foundation.

	Rotated Box Labeled Data	Horizontal Box Labeled Data	Single Point Labeled Data	10% 20% 30% Partially Labeled		
	RBox	HBox	Point	Partial	Price	Cost/Speed
RBox-supervised	- Rol-Transformer, KLD, GWD, PSC, ReDet, Oriented R-CNN, ...					
	😊	😞	😞	😞	\$86/1k	😞
HBox-supervised	- H2RBox, H2RBox-v2, OAOD, KCR, ...					
	😞	😊	😞	😞	\$63/1k	😞
Point-supervised	- PointOBB-v1/v2/v3, Point2RBox-v1/v2, ...					
	😞	😞	😊	😞	\$42/1k	😊
Semi-supervised	- Mean Teacher, Dense Teacher, SOOD, PseCo, MCL, ARSL, ...					
	😊	😞	😞	😊	\$17/0.2k \$0/0.8k	😊
Partially Weakly-supervised	- PWOOD (Ours)					
	😊	😊	😊	😊	\$8/0.2k \$0/0.8k	😊
	☑️ 20% HBoxes + 80% Unlabeled: 71.74 AP ₅₀		☑️ 20% Points + 80% Unlabeled: 53.59 AP ₅₀			

Figure 1. The main paradigmatic types of existing oriented object detection. Compared to these settings, our proposed Partial Weakly-supervised PWOOD offers high efficiency, low costs, and good performance (accuracy on DOTA-v1.0 is shown).

annotations has been dominant, as shown in Figure 1(a). However, obtaining rotated box annotations for a large number of images is extremely costly and labor-intensive, which poses a significant challenge for model training.

To address the issue of the difficulty in obtaining a large amount of data, semi-supervised oriented object detection (SOOD) [11, 35] algorithms have been proposed. As shown in Figure 1(d), these algorithms suggest using only a small amount of labeled data within the dataset and making full use of unlabeled data through the application of a self-training framework [14, 21, 22, 37], thereby enhancing the performance of the detector under the condition of small-batch labeled data. Moreover, weakly supervised oriented object detection (WOOD) [25, 46, 52, 53, 55] offers another cost-effective solution. Notably, methods such as H2RBox-

v2 [52], Point2RBox-v2 [54], and Wholly-WOOD [55] have demonstrated the feasibility of training detectors using horizontal bounding box annotations and single point annotations, as illustrated in Figure 1(b-c).

To further reduce annotation costs and fully leverage weakly annotated and unlabeled data, we propose a new oriented object detection framework, named Partial Weakly-Supervised Oriented Object Detection (PWOOD), that utilizes only a subset of weakly annotated data (e.g. horizontal bounding box or single point), as demonstrated in Figure 1(e). Inheriting the teacher-student paradigm [31], we introduce orientation learning and scale learning strategies to address the limitation of the teacher-student framework in learning orientation and scale information from partially weak annotations. This enables the student model to acquire the ability to learn the precise pose of the object, thereby achieving an Orientation-and-Scale-aware Student (OS-Student) model. However, a significant limitation of teacher-student paradigm is their reliance on static thresholds for pseudo-label selection [21, 35]. This approach often leads to threshold inconsistency, which can adversely affect the robustness and generalization ability of the model [37, 38, 60]. To tackle this challenge, we design a Class-Agnostic Pseudo-Label Filtering (CPF) based on a Gaussian Mixture Model. By employing maximum likelihood estimation, we dynamically adjust the filtering threshold, allowing the model to adaptively generate pseudo-labels that are more stable and aligned with the teacher’s performance. We apply this framework to a setting with partially annotated horizontal bounding boxes, achieving comparable performance to the traditional semi-supervised baseline. Additionally, we also validate the framework on datasets with partial point annotations.

Inspired by the teacher-student paradigm, we leverage a small amount of weakly-annotated data for pre-training, enabling the teacher model to generate pseudo-labels for unlabeled data. These pseudo-labels are then utilized to train the student model, allowing the student to learn from both the limited weakly-annotated data and abundant unlabeled data. As a result, the quality of pseudo-labels and the strategy employed for filtering them are pivotal to the overall performance of the model. However, a significant limitation of existing methods is their reliance on static thresholds for pseudo-label selection [21, 35]. This approach often leads to threshold inconsistency, which can adversely affect the robustness and generalization ability of the model [4, 37, 38, 60]. To tackle this challenge, we focus on enhancing the quality of pseudo-labels produced by the teacher model. We design a Class-Agnostic Pseudo-Label Filtering (CPF) based on a Gaussian Mixture Model. By employing maximum likelihood estimation, we dynamically adjust the filtering threshold, allowing the model to adaptively generate pseudo-labels that are more stable and

aligned with the teacher’s performance. This approach not only mitigates the issue of threshold inconsistency but also improves the model’s ability to handle diverse and complex scenarios, ultimately leading to more robust detection performance.

The contributions of this work are as follows:

- To our best knowledge, we propose the first Partial Weakly-Supervised Oriented Object Detection (PWOOD) framework, aiming to achieve competitive performance in a cost-conscious setting.
- We utilize partially weak annotations to enable the student model to learn orientation and scale information, resulting in an orientation-and-scale-aware student.
- Class-Agnostic Pseudo-Label Filtering is introduced for teacher models to address threshold inconsistency and reduce sensitivity to static thresholds, thereby enhancing the robustness of the model.
- Extensive experiments on DOTA-v1.0/v1.5/v2.0 and DIOR datasets show that our framework exhibits performance comparable to SOOD algorithms that rely on partially annotated rotated bounding boxes, thereby achieving superior results with weaker supervision.
- We generalize the PWOOD framework to support various annotation forms. Preliminary experiments show it reduces gaps between different annotations, offering cost-effective training options for detectors.

2. Related Work

Semi-supervised Oriented Object Detection: In semi-supervised detection, the teacher-student framework is a common paradigm [15, 20, 26, 29], where the teacher model generates pseudo-labels from unlabeled data to supervise the training of the student model, and the student model updates the teacher model using exponential moving average (EMA). For instance, SOOD [11] leverages optimal transport theory to define a cost matrix, calculating the distance between pseudo-labels and predictions. Meanwhile, MCL [35] builds upon the teacher-student paradigm by introducing Gaussian centeredness into the label assignment strategy and incorporating adaptive label assignment for unsupervised learning based on the characteristics of different feature layers. However, both models rely on rotated bounding box annotations for training, which are difficult to obtain in practice and come with high annotation costs, significantly increasing the overall training cost.

HBox-supervised Oriented Object Detection: Weakly supervised detection algorithms [2, 12, 42, 56, 62] aim to train detectors using more accessible and cost-effective annotations. Among them, numerous methods are based on horizontal bounding box annotations [17, 30, 33, 36, 62]. For instance, H2RBox [46] leverages horizontally annotated data and employs angle consistency loss, enabling the detector to learn orientation information. Based on

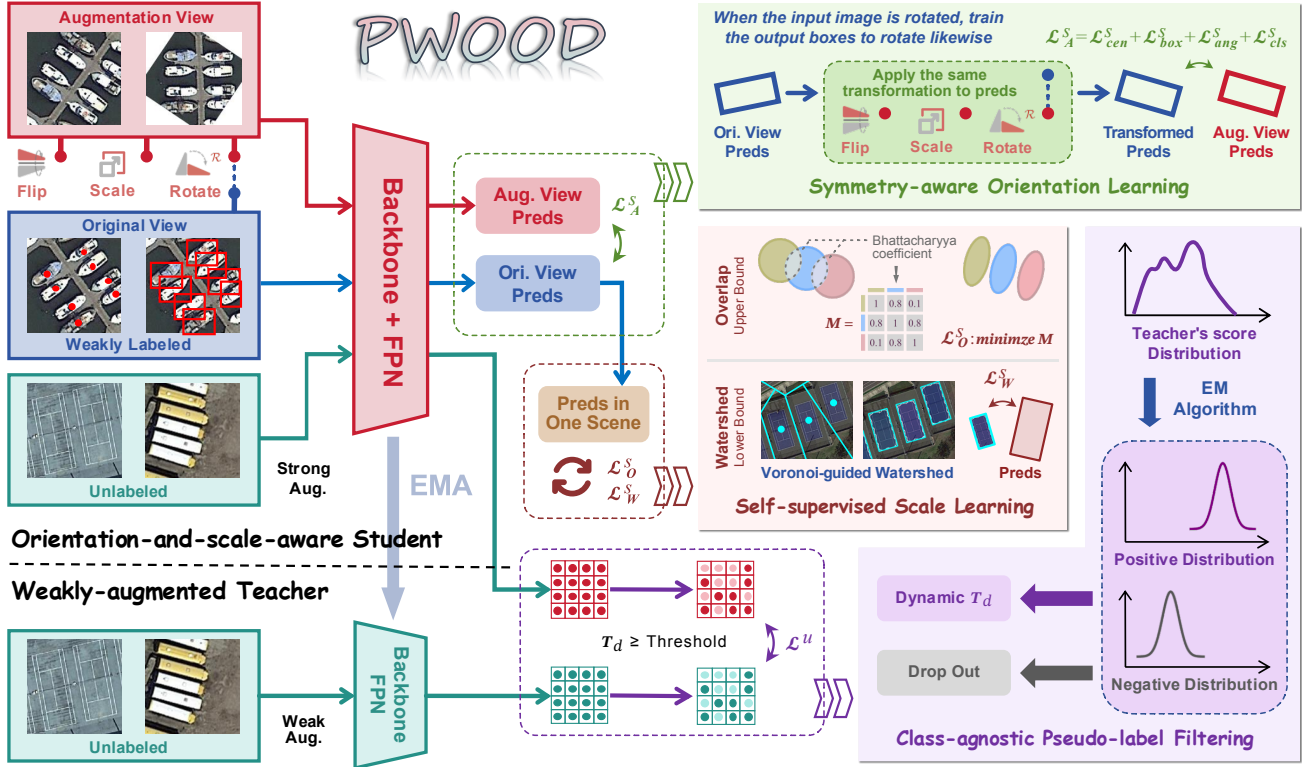


Figure 2. The overview of the proposed PWOOD framework. Orientation Learning and Scale Learning modules enables the Orientation-and-Scale-aware Student to learn both scale and orientation information from weakly annotated data, as well as Class-Agnostic Pseudo-Label Filtering mechanism enhances pseudo-label quality by leveraging dynamic thresholds.

H2RBox, H2RBox-v2 [52] introduces symmetric learning. In addition to random rotation augmentation, it incorporates vertical flipping and self-supervised symmetry learning, enriching the learning pathways for orientation information. Moreover, under horizontal bounding box supervision, AFWS [24] proposes an angle-free approach that decouples horizontal information from rotational information using concentric circles, simplifying the training process.

Point-supervised Oriented Object Detection: In recent studies, multiple approaches have been developed for oriented object detection using point supervision [5, 6, 10, 49]. For instance, P2RBox [3], PMHO [58] utilize SAM’s [13] zero-shot point-to-mask functionality to perform detection with point-based prompts. Another method, Point2RBox [53], adopts an end-to-end strategy by integrating diverse knowledge sources. Additionally, PointOBB [25] introduces a technique for generating rotated bounding boxes from points, employing scale-sensitive consistency and multiple instance learning. Building on this, PointOBB-v2 [54] enhances the process by constructing a class probability map and applying principal component analysis to produce pseudo RBox annotations, further pushing the boundaries of point-supervised detection. Point2RBox-v2 [54] incorporates novel losses based on Gaussian overlap and Voronoi tessellation to en-

force spatial layout constraints, along with additional modules such as edge loss, consistency loss, and copy-paste augmentation to further enhance its effectiveness.

3. Method

In this section, we elaborate on how to leverage orientation learning and scale learning strategies to form an Orientation-and-Scale-aware Student (OS-Student) and construct the PWOOD framework, achieving oriented object detection under the supervision of a small amount of weakly annotated data. First, we introduce our proposed PWOOD framework and its working mechanism. Next, we present the OS-Student model trained with scale learning and orientation learning modules. Additionally, to address the inconsistency in pseudo-label assignment thresholds for semi-supervised learning, we propose a Class-Agnostic Pseudo-Label Filtering (CPF). It ensures more reliable pseudo-label generation and enhances the overall robustness of the model. Finally, we delineate the composition of the overall loss function for the proposed model.

3.1. PWOOD Framework

Given the limited scale of weakly-annotated data, to fully exploit the potential of unlabeled data and further enhance

the performance of the OS-Student model, we utilize the teacher model to generate pseudo-labels [28] for unlabeled data. As illustrated in Figure 2, both the teacher and the OS-Student share identical architectures in their backbone, neck, and head components [4, 31]. During the pre-training phase, the weakly-annotated data is utilized to train the OS-Student. The OS-Student automatically learns the scale and orientation information from a small amount of weakly annotated data through the Orientation Learning and Scale Learning modules depicted in Figure 2. When the model reaches the burn-in step, the weights of the OS-Student are mirrored to the teacher. Subsequently, unlabeled data is introduced, and the model training enters the burn-in stage.

During the burn-in stage, the data undergoes weak and strong augmentation before being fed into the teacher and OS-Student networks [11], respectively, to obtain predictions. Based on the teacher’s confidence scores for each pseudo-box and a Gaussian Mixture Model, we employ a Class-Agnostic Pseudo-Label Filtering to filter the pseudo-boxes, thereby generating high-quality pseudo-labels. This approach enhances the robustness and accuracy of the model in leveraging both weakly-annotated and unlabeled data for improved object detection performance.

During this process, the student model is not only trained using pseudo-labels but also updates the weights of the teacher model through an Exponential Moving Average (EMA) [31] approach. This dynamic weight update mechanism allows the orientation-and-scale-aware capabilities learned by the student to be effectively transferred back to the teacher model, creating a positive feedback loop. As training progresses, the accuracy of the pseudo-labels generated by the teacher model gradually improves, further enhancing the ability of the OS-Student.

3.2. OS-Student

Orientation Learning: To enable the student model to learn orientation information, we introduce symmetry learning [52]. During training, the input images are vertically flipped or randomly rotated to generate transformed views. These views are fed into the network to obtain predictions for both the original and transformed images. The labeled data undergoes the same transformations, forming weakly-supervised pairs. Additionally, since there is a deterministic mapping relationship between the original image and its transformed views, the predictions of the model for the original and transformed views should satisfy the same mapping relationship, provided the prediction accuracy is high. Based on this principle, the predictions of the network for the original and transformed views form self-supervised pairs. Through the self-supervised and weakly supervised branches, the student model gains the ability to learn orientation information in a weakly annotated setting. We formulate an angle loss $\mathcal{L}_{\text{Ang}}^s$ to ensure that the OS-

Student can effectively learn orientation information from weakly annotated horizontal bounding boxes:

$$\mathcal{L}_{\text{Ang}}^s = \begin{cases} L_{\text{Ang}}^s(\theta_{\text{flip}} + \theta, 0) & \text{trans} = \text{flip} \\ L_{\text{Ang}}^s(\theta_{\text{rot}} - \theta, \mathcal{R}) & \text{trans} = \text{rotate}(\theta). \end{cases} \quad (1)$$

The computation of the angle loss is related to the image transformation method *trans*, which involves either vertical flipping or random rotation by an angle θ . L_{Ang}^s denotes Smooth-L1 loss.

Scale Learning: Considering the presence of even weaker annotation forms in the dataset, such as single point annotations, which lack scale information, we introduce a scale learning strategy to enable the student model to learn scale information effectively under such weakly annotated conditions. We employ spatial layout learning [54] to guide the model’s scale prediction accuracy by estimating upper and lower bounds on object scales.

To obtain the upper bound of the object scale, we treat the oriented bounding box as a Gaussian distribution and introduce the Bhattacharyya coefficient [47] to measure the overlap between Gaussian distributions. We find the upper bound by minimizing the Gaussian overlap between different predicted bounding boxes. Consequently, we derive the Gaussian overlap loss as follows:

$$\mathcal{L}_O^s = \frac{1}{N} \sum_{i,j=1, i \neq j}^N B(\mathcal{N}_i, \mathcal{N}_j), \quad (2)$$

where N denotes the number of predicted bounding boxes, \mathcal{N}_i is the Gaussian distribution of the i -th box, and $B(\mathcal{N}_i, \mathcal{N}_j)$ denotes the Bhattacharyya coefficient between the i -th and j -th predicted bounding boxes.

Moreover, to obtain the lower bound of the object scale, the Voronoi diagram [1] and the watershed algorithm [34] are introduced to calculate Voronoi Watershed Loss. Using the ridges of the Voronoi diagram as background markers and point annotations as foreground markers, we apply the watershed algorithm to segment the image and obtain the basin regions for each object. By rotating these watershed regions to align with the current predicted orientation, we calculate the regression objects for width and height. Finally, we use the Gaussian Wasserstein Distance loss [47] to regress the width w_t and height h_t of the objects:

$$\mathcal{L}_W^s = L_{\text{GWD}} \left(\begin{bmatrix} w/2 & 0 \\ 0 & h/2 \end{bmatrix}^2, \begin{bmatrix} w_t/2 & 0 \\ 0 & h_t/2 \end{bmatrix}^2 \right). \quad (3)$$

And the supervised loss \mathcal{L}^s of OS-Student is as follows:

$$\begin{aligned} \mathcal{L}^s = & \alpha_1 \mathcal{L}_{\text{cls}}^s(p_{(x,y)}, c_{(x,y)}) \\ & + \alpha_2 \mathcal{L}_{\text{cen}}^s(\text{cen}'_{(x,y)}, \text{cen}_{(x,y)}) \\ & + \alpha_3 \mathcal{L}_{\text{box}}^s(\text{pr}_{(x,y)}, \text{gt}_{(x,y)}) \\ & + \alpha_4 \mathcal{L}_{\text{Ang}}^s + \alpha_5 \mathcal{L}_O^s + \alpha_6 \mathcal{L}_W^s, \end{aligned} \quad (4)$$

where class loss \mathcal{L}_{cls}^s , centerness loss \mathcal{L}_{cen}^s , and box loss \mathcal{L}_{box}^s are inherited from the backbone detector.

According to FCOS [32], each positive point (x, y) on the feature map corresponds to a potential object center or anchor point in the input image. p, cen', pr represent the category score, centerness, and rotated bounding box predictions of the OS-Student, while c, cen, gt are the corresponding weak ground truth labels. $\mathcal{L}_{cls}^s, \mathcal{L}_{cen}^s$ and \mathcal{L}_{box}^s denotes focal loss [19], cross entropy loss, and IoU loss [50], respectively. For hyperparameters, $\alpha_1, \alpha_2, \alpha_3$ are all set to 1, following FCOS[32] where this configuration was shown effective. $(\alpha_4, \alpha_5, \alpha_6)$ is set to $(0.2, 10, 5)$ by default, ablations are provided in the supplementary materials.

3.3. Class-Agnostic Pseudo-Label Filtering

The quality of pseudo-labels generated by the teacher model is a critical factor influencing the overall performance of the detection framework. Most existing approaches [21, 28, 35, 41] adopt a static thresholding strategy to filter pseudo-labels, which is typically determined empirically and lacks adaptability to the inherent characteristics of the data and the dynamic distribution of pseudo-labels across different training stages. Specifically, during the initial phases of training, the teacher model tends to produce pseudo-labels with relatively low confidence scores due to its underdeveloped representation capability, while as training progresses, the teacher model gradually improves, leading to higher-quality pseudo-labels with more reliable confidence estimates. And we find that during the training stage, the model exhibits high sensitivity to threshold settings.

The issue above highlights the need for a more adaptive and data-driven approach [38, 60] to pseudo-label selection in semi-supervised object detection. To address this, we propose Class-Agnostic Pseudo-Label Filtering (CPF). Based on a Gaussian Mixture Model [37, 59], we model the scores of pseudo boxes generated by the teacher as a mixed model $\mathcal{P}(s)$ of two one-dimensional Gaussian distributions:

$$\mathcal{P}(s) = w_p \mathcal{N}_p(\mu_p, (\sigma_p)^2) + w_n \mathcal{N}_n(\mu_n, (\sigma_n)^2), \quad (5)$$

where $\mathcal{N}_p, \mathcal{N}_n$ denote the distributions of positive samples and negative samples, respectively. The initial mean of the positive sample distribution $\mu_p^{(0)}$ is set to the maximum of the predicted score s , while the negative value $\mu_n^{(0)}$ is set to the lowest predicted score. The weights of the two one-dimensional Gaussian distributions $w_p^{(0)}$ and $w_n^{(0)}$ in the mixture model are both initialized to 0.5. Then, we employ the expectation-maximization (EM) algorithm to deduce the posterior \mathcal{P}_p , which represents the likelihood that a detection ought to be designated as the pseudo-object for the student:

$$T_d = \underset{s}{\operatorname{argmax}} \mathcal{P}_p(s, \mu_p, (\sigma_p)^2). \quad (6)$$

3.4. Overall Loss

In our proposed PWOOD framework, the overall loss function of the model consists of two distinct components: the supervised loss \mathcal{L}^s within the OS-Student and the unsupervised loss \mathcal{L}^u derived from the pseudo-labels generated by the teacher model to guide the learning of the OS-Student. We elaborate on the former in Section 3.2, and the latter is defined as follows:

$$\mathcal{L}^u = \omega (\mathcal{L}_{cls}^u(\mathcal{T}_{(x,y)}^c, \mathcal{S}_{(x,y)}^c) + \mathcal{L}_{cen}^u(\mathcal{T}_{(x,y)}^{cen}, \mathcal{S}_{(x,y)}^{cen}) + \mathcal{L}_{box}^u(\mathcal{T}_{(x,y)}^{logit}, \mathcal{S}_{(x,y)}^{logit})). \quad (7)$$

The predictions of teacher and student models, denoted as \mathcal{T} and \mathcal{S} , are represented by the tuple $(c, cen, logit)$ where c is the class score indicating the confidence of the predicted class, cen is the predicted center-ness measuring how close the point is to the center of the object, $logit$ represents the distances from the point to the predicted bounding box's left, top, right, and bottom boundaries. \mathcal{L}_{cls}^u and \mathcal{L}_{cen}^u are binary cross-entropy loss for classification and center-ness prediction, respectively. \mathcal{L}_{box}^u is the Smooth L1 loss, applied to the regression of the bounding box distances. The weight ω of each loss component is associated with the score of each point, ensuring that points with higher confidence contribute more to the overall loss. This weighting mechanism helps the model focus on high-quality predictions, improving both localization and classification accuracy. The overall loss of PWOOD is as follows:

$$\mathcal{L} = \alpha \mathcal{L}^s + \beta \mathcal{L}^u, \quad (8)$$

both α and β are set to 1, ablations are in the supplement.

4. Experiment

Our experiments are carried out on the orientation detection tool kits: MMRotate 0.3.4 [61] and Pytorch 1.13.1 [27].

4.1. Datasets and Experimental Setting

DOTA Dataset-Partial: We conducted experiments on all three versions of the DOTA dataset [39]. DOTA-v1.0 contains 188,282 instances across 15 categories, while DOTA-v1.5/2.0 adds numerous small object annotations. Following the SOOD [11] method, we randomly sampled 10%, 20%, and 30% of the images from the DOTA training set to serve as annotated data. During the data loading process, these annotated images underwent specific transformations to remove orientation and scale information, resulting in semi-weakly annotated data. The remaining images are treated as unannotated data. For validation purposes, the DOTA-v1.5 validation set is employed.

DIOR Dataset-Partial: DIOR [16] comprises a total of 23,463 images, with 11,725 images allocated for training and 11,738 images for testing. To facilitate our experiments,

Table 1. mAP comparison on DIOR *val set*, DOTA-v1.0 *val set* and DOTA-v2.0 *val set*.

Task	Method	DIOR			DOTA-v1.0			DOTA-v2.0		
		10%	20%	30%	10%	20%	30%	10%	20%	30%
WOOD	H2RBox-v2 [52]	44.54	51.33	53.45	47.96	54.38	58.65	19.07	28.56	31.73
	Point2RBox-v2 [54]	28.77	32.81	33.07	35.24	40.39	45.09	12.18	16.74	16.62
SOOD	Vanilla Baseline (w/ Partial RBox)	54.01	57.07	60.25	56.03	62.82	64.97	24.77	34.03	37.30
PWOOD (ours)	w/ Partial HBox Annotations	54.33	57.89	60.42	56.92	62.93	65.42	31.03	36.39	40.27
	w/ Partial Point Annotations	32.04	35.17	36.44	42.35	45.01	49.12	13.44	18.49	23.85

Table 2. mAP comparison on DOTA-v1.5 *val set*.

Task	Method	DOTA-v1.5 Dataset-Partial		
		10%	20%	30%
WOOD	H2RBox-v2	42.19	49.01	55.19
	Point2RBox-v2	32.69	36.03	38.30
SOOD (w/ Partial RBox)	SOOD	48.63	55.58	59.23
	Dense Teacher	46.90	53.93	57.86
	PseCo	48.04	55.28	58.03
	ARSL	48.17	55.34	59.02
	SOOD++	50.48	57.44	61.51
	MCL	52.61	59.63	62.63
	Vanilla Baseline	49.53	58.28	61.00
PWOOD (ours)	w/ Partial HBox	52.87	59.36	61.58
	w/ Partial Point	35.33	41.54	43.02

we have offline partitioned the DIOR training set into subsets containing 10%, 20%, and 30% of the data as supervised datasets, while the remaining data is treated as unsupervised. The DIOR test set is utilized for both validation and testing purposes, ensuring a comprehensive evaluation of the model’s performance.

Experimental Setting: We adopt the FCOS [32] detector with ResNet50 [9] backbone and FPN [18] neck. We employ a simplified version of the MCL¹, excluding the GCA and CCSL modules [35], as our SOOD baseline, referred to as the Vanilla Baseline. In addition, we further compare the performance of our proposed PWOOD framework with weakly supervised detectors [52, 54] trained on partially weakly annotated data to ensure the comprehensiveness and integrity of experiments. To ensure a fair and comprehensive evaluation, we adopt Average Precision (AP) as the primary metric for benchmarking against existing literature. In the experiments, PWOOD is trained for 180k iterations under the 30% partial and full settings, while for the 10% and 20% partial settings, it is trained for 120k iterations [35]. All models are trained using the AdamW optimizer [23].

4.2. Main Results

Partial Horizontal Box: As illustrated in Table 2, our proposed PWOOD framework demonstrates superior performance on the DOTA-v1.5 Dataset-Partial with 10%, 20%,

¹Some of MCL’s components lack universality. For instance, GCA cannot be deployed in Point2Rbox-v2, which is independent of centerness.

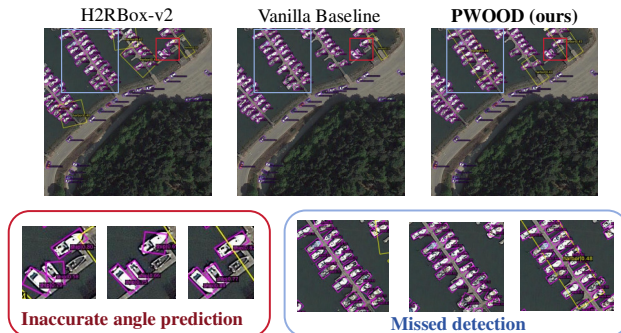


Figure 3. Visualized performance comparison of PWOOD with H2RBox-v2 and the Vanilla Baseline.

and 30% partial horizontal bounding box annotations, outperforming the SOOD baseline methods that utilize corresponding proportions of rotated bounding box annotations. Specifically, under 10%, 20%, and 30% HBox annotations, PWOOD improves mAP by 3.34%, 1.08%, and 0.58%, respectively, compared to the SOOD baseline method. It means that PWOOD can achieve comparable or even better performance at a lower cost. On the DIOR Dataset-Partial in Table 1, even when using weakly annotated horizontal boxes, PWOOD achieves performance comparable to the Vanilla Baseline trained with the same proportion of costly rotated box annotations. This further validates that PWOOD delivers equivalent results (54.33% vs. 54.01%, 57.89% vs. 57.07%, 60.42% vs. 60.25%) at a significantly lower cost compared to high-cost training methods.

Furthermore, we compare our approach with the weakly supervised algorithm H2RBox-v2 [52], which only employs partial horizontal bounding box annotations during training. Our model significantly outperforms H2RBox-v2 in the partially weak annotation setting. Specifically, on the DOTA-v1.5 Dataset-Partial, PWOOD achieves large margins of improvement, with gains of 10.68%, 10.35%, and 6.39% for the 10%, 20%, and 30% annotation ratios, respectively. Similarly, on the DIOR Dataset-Partial in Table 1, PWOOD demonstrates mAP improvements ranging from 6.56% to 9.79%. These results emphasize that PWOOD can effectively mine valid information from unlabeled data, highlighting its efficiency and effectiveness in utilizing limited annotation resources.

Table 3. Detection performance of each category on the DOTA-v1.0 *test set* and the mean AP₅₀ of all categories.

Methods	PL ¹	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	AP ₅₀
▼ <i>RBox-supervised OOD</i>																
RepPoints (2019) [48]	86.7	81.1	41.6	62.0	76.2	56.3	75.7	90.7	80.8	85.3	63.3	66.6	59.1	67.6	33.7	68.45
RetinaNet (2017) [19]	88.2	77.0	45.0	69.4	71.5	59.0	74.5	90.8	84.9	79.3	57.3	64.7	62.7	66.5	39.6	68.69
GWD (2021) [45]	89.3	75.4	47.8	61.9	79.5	73.8	86.1	90.9	84.5	79.4	55.9	59.7	63.2	71.0	45.4	71.66
FCOS (2019) [32]	89.1	76.9	50.1	63.2	79.8	79.8	87.1	90.4	80.8	84.6	59.7	66.3	65.8	71.3	41.7	72.44
S ² A-Net (2022) [8]	89.2	83.0	52.5	74.6	78.8	79.2	87.5	90.9	84.9	84.8	61.9	68.0	70.7	71.4	59.8	75.81
▼ <i>HBox-supervised OOD</i>																
H2RBox (2023) [46]	88.5	73.5	40.8	56.9	77.5	65.4	77.9	90.9	83.2	85.3	55.3	62.9	52.4	63.6	43.3	67.82
EIE-Det (2024) [36]	87.7	70.2	41.5	60.5	80.7	76.3	86.3	90.9	82.6	84.7	53.1	64.5	58.1	70.4	43.8	70.10
H2RBox-v2 (2023) [52] ²	88.9	70.7	47.1	60.5	79.7	73.6	87.4	90.9	82.3	75.5	60.3	64.3	64.1	68.4	40.7	70.30
H2RBox-v2 (2023) [52]	89.1	74.6	47.5	59.8	80.7	73.1	87.8	90.9	85.6	74.1	60.7	61.7	65.8	71.8	56.0	71.96
▼ <i>Point-supervised OOD</i>																
Point2RBox (2024) [53]	62.9	64.3	14.4	35.0	28.2	38.9	33.3	25.2	2.2	44.5	3.4	48.1	25.9	45.0	22.6	34.07
Point2RBox+SK (2024) [53]	53.3	63.9	3.7	50.9	40.0	39.2	45.7	76.7	10.5	56.1	5.4	49.5	24.2	51.2	33.8	40.27
PointOBB-v3 (2025) [57]	30.9	39.4	13.5	22.7	61.2	7.0	43.1	62.4	59.8	47.3	2.7	45.1	16.8	55.2	11.4	41.29
Point2RBox-v2 (2025) [54] ²	77.9	51.6	7.5	35.3	69.6	58.5	75.1	88.3	57.6	73.1	12.3	34.1	29.6	47.2	17.8	49.04
Point2RBox-v2 (2025) [54]	78.4	52.7	8.3	40.9	71.0	60.5	74.7	88.7	65.5	72.1	24.4	26.1	30.1	50.7	21.0	51.00
▼ <i>Semi-supervised OOD</i>																
MCL (2025) [35] [†]	88.5	79.6	46.0	65.1	80.4	81.9	87.7	90.9	78.5	85.6	57.0	68.3	66.5	74.1	54.6	73.64
Vanilla Baseline [†]	88.5	72.9	42.4	56.3	78.4	80.3	87.2	90.8	78.3	84.1	55.4	59.8	70.7	73.7	43.3	70.80
▼ <i>Partial Weakly-supervised OOD (ours)</i>																
w/ Partial Horizontal Box [†]	89.0	69.3	49.7	50.6	79.3	74.4	86.2	90.9	81.9	85.2	56.2	65.1	69.1	75.8	53.4	71.74
w/ Partial Single Point [†]	79.0	60.8	14.6	38.5	70.9	66.5	73.5	86.5	69.8	74.3	29.1	13.6	33.7	59.2	33.8	53.59

¹PL: Plane, BD: Baseball diamond, BR: Bridge, GTF: Ground track field, SV: Small vehicle, LV: Large vehicle, SH: Ship, TC: Tennis court,

BC: Basketball court, ST: Storage tank, SBF: Soccer-ball field, RA: Roundabout, HA: Harbor, SP: Swimming pool, HC: Helicopter.

²Only the training set is used for training. Others without superscripts are trained using the trainval set by default, as in the corresponding published papers.

[†]The fully/weakly labeled training data is the training set, and the unlabeled training data is the validation set.

Table 4. Ablation with different levels of noise adding to HBox/Point annotations on DOTA-v1.0 *val set* and DOTA-v1.5 *val set*. Noise sampled from a uniform distribution $[-\sigma H, +\sigma H]$, where H represents the height of the objects, is added to annotations.

Dataset	Noise	WOOD		PWOOD	
		Point2RBox-v2	H2RBox-v2	w/ Partial Point	w/ Partial HBox
DOTA-v1.0	0%	40.39	54.38	45.01	62.93
	10%	30.41	50.55	46.34	59.88
	30%	24.27	41.75	35.50	55.32
DOTA-v1.5	0%	36.03	49.01	41.54	59.36
	10%	24.23	44.52	40.81	56.07
	30%	21.92	35.87	31.67	51.49

Partial Single Point: Given the diversity of weak annotation forms, to validate the robustness of the PWOOD framework under partial weakly supervised settings, we further evaluated its performance in a setting with partial single point. Experimental results in Table 2 demonstrate that, upon the elimination of scale-related information, the performance of PWOOD experienced a decline; however, when compared to the weakly supervised algorithm with partial single point, Point2RBox-v2 [54], our PWOOD still demonstrated significant superiority. Specifically, on the DOTA-v1.5 Dataset-Partial with 10%, 20%, and 30% partial single point, PWOOD achieved improvements in mAP by 2.64%, 5.51%, and 4.72%, respectively. Under the same annotation ratios on the DIOR Dataset-Partial, it also showed an increase in mAP ranging from 2.36% to 3.37%. This demonstrates that our framework exhibits universality across different forms of partially weak annotations.

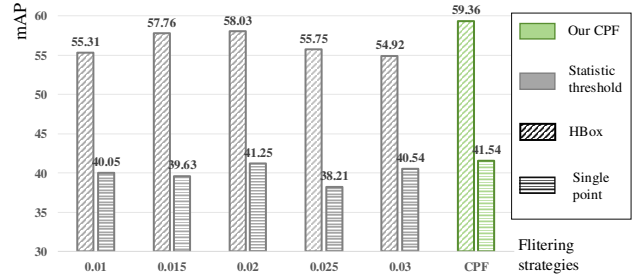


Figure 4. Ablation of pseudo-label filtering strategies.

More Results: We also conducted experiments on DOTA-v1.0/v2.0. As shown in Table 1, PWOOD demonstrates significant performance improvements over WOOD under three different HBox and Point annotation ratios, further proving that PWOOD effectively leverages unlabeled data. Specifically, PWOOD outperforms the Vanilla Baseline

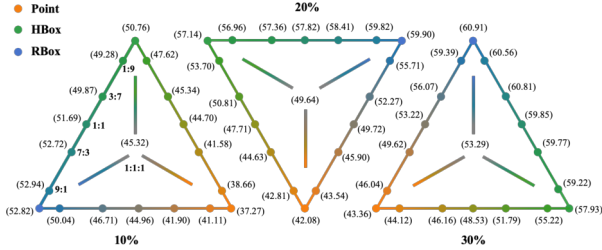


Figure 5. The preliminary experimental results of joint training with various annotations on DOTA-v1.5 Dataset-Partial

across DOTA-v1.0/v1.5/v2.0 datasets under all HBox annotation ratios. Moreover, As detection difficulty increases, the relative mAP gain of PWOOD over Vanilla Baseline becomes more pronounced. On DOTA-v2.0 which includes more small objects, PWOOD achieves mAP improvements of 6.26%, 2.36%, and 2.97%, respectively, compared to the gain on DOTA-v1.0 (0.89%, 0.11%, and 0.45, respectively). This suggests that PWOOD exhibits unique advantages in complex scenes with small objects.

What’s more, we evaluate the performance of WOOD, SOOD, and PWOOD on the DOTA-v1.0 *test set* at full scale. Both WOOD, SOOD, and PWOOD are trained on the DOTA-v1.0 *train set*, while WOOD and PWOOD utilize the DOTA-v1.0 *val set* as unlabeled data. As shown in Table 3, compared to WOOD, PWOOD achieves improvements of 1.44% (71.74% vs. 70.30%) and 4.55% (53.59% vs. 49.04%) under the weak annotation settings of partial HBox and partial single point, respectively, proving that PWOOD fully exploits the potential of unlabeled data. Additionally, compared to the Vanilla Baseline, PWOOD achieves comparable performance, indicating that PWOOD can deliver highly competitive results using low-cost annotations. As illustrated in Figure 3, the visual comparison among the three frameworks shows that PWOOD exhibits fewer instances of inaccurate angle predictions and missed detections compared to the other two methods. This further validates that PWOOD not only enhances accuracy but also improves robustness in handling complex scenarios.

4.3. Ablation Study

Threshold Sensitivity Analysis: The selection threshold for pseudo-labels is critical to their quality. However, static thresholds often fail to adapt to the specific characteristics of different datasets and training stages, resulting in significant sensitivity of the model to static thresholds. To demonstrate this, we conducted multiple experiments with static thresholds on the DOTA-v1.5 Dataset-Partial with 20% weak annotations, as illustrated in Figure 4. It is evident that the model’s performance varies considerably under different thresholds. Specifically, a slight change in the static threshold from 0.02 to 0.03 leads to a 3.09% drop in mAP. Furthermore, as shown in Figure 4, the best-

performing static threshold achieves a validation mAP of 58.03% and 41.25%, respectively, while the introduction of CPF boosts the mAP to 59.36% and 41.54%, respectively. This indicates that CPF mechanism effectively improves pseudo-label quality and enhances model performance.

Robustness Analysis Against Noise: As shown in Table 4, by comparing the performance of WOOD and our PWOOD framework under varying levels of noise interference on the DOTA-v1.5 and DOTA-v1.0 datasets, it is evident that PWOOD exhibits superior robustness compared to WOOD. For example, when 10% and 30% noise is added to the DOTA-v1.0 dataset, under 20% horizontal box supervision, H2RBox-v2 shows mAP drops of 3.83% and 12.63%, respectively, while PWOOD only experiences drops of 3.05% and 7.61%. This indicates that PWOOD not only achieves superior performance compared to WOOD but also reduces performance degradation under noise interference.

4.4. More Discussion

Building upon these demonstrated achievements and considering the inherent diversity of annotation standards in practical applications, we identify a significant opportunity to enhance the PWOOD framework’s versatility. Specifically, enabling the framework to support joint training with multi-format labeled data (RBox, HBox, and Point) emerges as a highly promising direction for substantially reducing the challenges associated with training data acquisition. The experimental results are shown in Figure 5. Notably, at 20% multi-format annotations, substituting 10% of RBox annotations with equivalent proportions of HBoxes results in merely 0.08% (59.90% vs. 59.82%) mAP degradation, respectively. These results demonstrate that our PWOOD successfully bridges the cost-performance gap and provides better cost-accuracy tradeoffs.

5. Conclusion

In this paper, we present a new framework, PWOOD, for oriented object detection using a weaker labeling paradigm to further reduce the annotation cost. Within this framework, we introduce the orientation and scale learning modules, endowing the student model with the capability to learn orientation and scale information and resulting in the OS-Student. Moreover, to mitigate the model’s sensitivity to static pseudo-label filtering thresholds, we propose CPF that dynamically filters pseudo-labels to enhance the model’s generalization capability. Our proposed PWOOD model reduces the price in both annotation speed and cost. Extensive experiments on benchmarks including DOTA-v1.0/v1.5/v2.0 and DIOR demonstrate that, whether using the partial horizontal box setting or the single point setting, PWOOD achieves performance comparable to, or even surpassing existing WOOD and SOOD algorithms while significantly reducing the annotation costs.

References

- [1] Franz Aurenhammer. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys*, 23(3):345–405, 1991. 4
- [2] Hakan Bilen, Marco Pedersoli, and Tinne Tuytelaars. Weakly supervised object detection with convex clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1081–1089, 2015. 2
- [3] Guangming Cao, Xuehui Yu, Wenwen Yu, Xumeng Han, Xue Yang, Guorong Li, Jianbin Jiao, and Zhenjun Han. P2rbox: Point prompt oriented object detection with sam. *arXiv preprint arXiv:2311.13128*, 2023. 3
- [4] Binghui Chen, Pengyu Li, Xiang Chen, Biao Wang, Lei Zhang, and Xian-Sheng Hua. Dense learning based semi-supervised object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4815–4824, 2022. 2, 4
- [5] Liangyu Chen, Tong Yang, Xiangyu Zhang, Wei Zhang, and Jian Sun. Points as queries: Weakly semi-supervised object detection by points. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8823–8832, 2021. 3
- [6] Pengfei Chen, Xuehui Yu, Xumeng Han, Najmul Hassan, Kai Wang, Jiachen Li, Jian Zhao, Humphrey Shi, Zhenjun Han, and Qixiang Ye. Point-to-box network for accurate object detection via single point supervision. In *European Conference on Computer Vision*, pages 51–67. Springer, 2022. 3
- [7] Jian Ding, Nan Xue, Yang Long, Gui-Song Xia, and Qikai Lu. Learning roi transformer for oriented object detection in aerial images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2849–2858, 2019. 1
- [8] Jiaming Han, Jian Ding, Jie Li, and Gui-Song Xia. Align deep features for oriented object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022. 7
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 6
- [10] Shitian He, Huanxin Zou, Yingqian Wang, Boyang Li, Xu Cao, and Ning Jing. Learning remote sensing object detection with single point supervision. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2023. 3
- [11] Wei Hua, Dingkang Liang, Jingyu Li, Xiaolong Liu, Zhikang Zou, Xiaoqing Ye, and Xiang Bai. Sood: Towards semi-supervised oriented object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15558–15567, 2023. 1, 2, 4, 5
- [12] Javed Iqbal, Muhammad Akhtar Munir, Arif Mahmood, Afshen Rafaqat Ali, and Mohsen Ali. Leveraging orientation for weakly supervised object detection with application to firearm localization. *Neurocomputing*, 440:310–320, 2021. 2
- [13] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023. 3
- [14] Gang Li, Xiang Li, Yujie Wang, Yichao Wu, Ding Liang, and Shanshan Zhang. Pseco: Pseudo labeling and consistency training for semi-supervised object detection. In *European Conference on Computer Vision*, pages 457–472. Springer, 2022. 1
- [15] Hengduo Li, Zuxuan Wu, Abhinav Shrivastava, and Larry S Davis. Rethinking pseudo labels for semi-supervised object detection. In *AAAI Conference on Artificial Intelligence*, pages 1314–1322, 2022. 2
- [16] Ke Li, Gang Wan, Gong Cheng, Liquu Meng, and Junwei Han. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159:296–307, 2020. 5
- [17] Wentong Li, Wenyu Liu, Jianke Zhu, Miaomiao Cui, Xian-Sheng Hua, and Lei Zhang. Box-supervised instance segmentation with level set evolution. In *European Conference on Computer Vision*, pages 1–18. Springer, 2022. 2
- [18] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2117–2125, 2017. 6
- [19] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *IEEE International Conference on Computer Vision*, pages 2980–2988, 2017. 5, 7
- [20] Liang Liu, Boshen Zhang, Jiangning Zhang, Wuhao Zhang, Zhenye Gan, Guanzhong Tian, Wenbing Zhu, Yabiao Wang, and Chengjie Wang. Mixteacher: Mining promising labels with mixed scale teacher for semi-supervised object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7370–7379, 2023. 2
- [21] Yen-Cheng Liu, Chih-Yao Ma, Zijian He, Chia-Wen Kuo, Kan Chen, Peizhao Zhang, Bichen Wu, Zsolt Kira, and Peter Vajda. Unbiased teacher for semi-supervised object detection. *arXiv preprint arXiv:2102.09480*, 2021. 1, 2, 5
- [22] Yen-Cheng Liu, Chih-Yao Ma, and Zsolt Kira. Unbiased teacher v2: Semi-supervised object detection for anchor-free and anchor-based detectors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9819–9828, 2022. 1
- [23] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 6
- [24] Junyan Lu, Qinglei Hu, Ruifei Zhu, Yali Wei, and Tie Li. Afws: Angle-free weakly-supervised rotating object detection for remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 3
- [25] Junwei Luo, Xue Yang, Yi Yu, Qingyun Li, Junchi Yan, and Yansheng Li. Pointobb: Learning oriented object detection via single point supervision. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16730–16740, 2024. 1, 3
- [26] Yuxiang Nie, Chaowei Fang, Lechao Cheng, Liang Lin, and Guanbin Li. Adapting object size variance and class imbalance for semi-supervised object detection. In *AAAI Conference on Artificial Intelligence*, pages 1966–1974, 2023. 2
- [27] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming

- Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32, 2019. 5
- [28] Kihyuk Sohn, Zizhao Zhang, Chun-Liang Li, Han Zhang, Chen-Yu Lee, and Tomas Pfister. A simple semi-supervised learning framework for object detection. *arXiv preprint arXiv:2005.04757*, 2020. 4, 5
- [29] Peize Sun, Yi Jiang, Enze Xie, Wenqi Shao, Zehuan Yuan, Changhu Wang, and Ping Luo. What makes for end-to-end object detection? In *International Conference on Machine Learning*, pages 9934–9944. PMLR, 2021. 2
- [30] Yongqing Sun, Jie Ran, Feng Yang, Chenqiang Gao, Takayuki Kurozumi, Hideaki Kimata, and Ziqi Ye. Oriented object detection for remote sensing images based on weakly supervised learning. In *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–6. IEEE, 2021. 2
- [31] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in Neural Information Processing Systems*, 30, 2017. 2, 4
- [32] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection. In *IEEE/CVF International Conference on Computer Vision*, pages 9627–9636, 2019. 5, 6, 7
- [33] Zhi Tian, Chunhua Shen, Xinlong Wang, and Hao Chen. Boxinst: High-performance instance segmentation with box annotations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5443–5452, 2021. 2
- [34] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, 1991. 4
- [35] Chenxu Wang, Chunyan Xu, Ziqi Gu, and Zhen Cui. Multi-clue consistency learning to bridge gaps between general and oriented object in semi-supervised detection. *arXiv preprint arXiv:2407.05909*, 2024. 1, 2, 5, 6, 7
- [36] Linfei Wang, Yibing Zhan, Xu Lin, Baosheng Yu, Liang Ding, Jianqing Zhu, and Dapeng Tao. Explicit and implicit box equivariance learning for weakly-supervised rotated object detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024. 2, 7
- [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 2023 ieee. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3240–3249, 2022. 1, 2, 5
- [38] Yidong Wang, Hao Chen, Qiang Heng, Wenxin Hou, Yue Fan, Zhen Wu, Jindong Wang, Marios Savvides, Takahiro Shinozaki, Bhiksha Raj, et al. Freematch: Self-adaptive thresholding for semi-supervised learning. *arXiv preprint arXiv:2205.07246*, 2022. 2, 5
- [39] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3974–3983, 2018. 5
- [40] Xingxing Xie, Gong Cheng, Jiabao Wang, Xiwen Yao, and Junwei Han. Oriented r-cnn for object detection. In *IEEE/CVF International Conference on Computer Vision*, pages 3520–3529, 2021. 1
- [41] Mengde Xu, Zheng Zhang, Han Hu, Jianfeng Wang, Lijuan Wang, Fangyun Wei, Xiang Bai, and Zicheng Liu. End-to-end semi-supervised object detection with soft teacher. In *IEEE/CVF International Conference on Computer Vision*, pages 3060–3069, 2021. 5
- [42] Ke Yang, Dongsheng Li, and Yong Dou. Towards precise end-to-end weakly supervised object detection network. In *IEEE/CVF International Conference on Computer Vision*, pages 8372–8381, 2019. 2
- [43] Xue Yang, Jirui Yang, Junchi Yan, Yue Zhang, Tengfei Zhang, Zhi Guo, Xian Sun, and Kun Fu. Scrdet: Towards more robust detection for small, cluttered and rotated objects. In *IEEE/CVF International Conference on Computer Vision*, pages 8232–8241, 2019. 1
- [44] Xue Yang, Junchi Yan, Ziming Feng, and Tao He. R3det: Refined single-stage detector with feature refinement for rotating object. In *AAAI Conference on Artificial Intelligence*, pages 3163–3171, 2021. 1
- [45] Xue Yang, Junchi Yan, Ming Qi, Wentao Wang, Xiaopeng Zhang, and Tian Qi. Rethinking rotated object detection with gaussian wasserstein distance loss. In *38th International Conference on Machine Learning*, pages 11830–11841, 2021. 7
- [46] Xue Yang, Gefan Zhang, Wentong Li, Xuehui Wang, Yue Zhou, and Junchi Yan. H2rbox: Horizontal box annotation is all you need for oriented object detection. *International Conference on Learning Representations*, 2023. 1, 2, 7
- [47] Xue Yang, Gefan Zhang, Xiaojiang Yang, Yue Zhou, Wentao Wang, Jin Tang, Tao He, and Junchi Yan. Detecting rotated objects as gaussian distributions and its 3-d generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4335–4354, 2023. 4
- [48] Ze Yang, Shaohui Liu, Han Hu, Liwei Wang, and Stephen Lin. Reppoints: Point set representation for object detection. In *IEEE/CVF International Conference on Computer Vision*, pages 9656–9665, 2019. 7
- [49] Xinyi Ying, Li Liu, Yingqian Wang, Ruoqing Li, Nuo Chen, Zaiping Lin, Weidong Sheng, and Shilin Zhou. Mapping degeneration meets label evolution: Learning infrared small target detection with single point supervision. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15528–15538, 2023. 3
- [50] Jiahui Yu, Yuning Jiang, Zhangyang Wang, Zhimin Cao, and Thomas Huang. Unitbox: An advanced object detection network. In *24th ACM International Conference on Multimedia*, pages 516–520, 2016. 5
- [51] Yi Yu and Feipeng Da. On boundary discontinuity in angle regression based arbitrary oriented object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–14, 2024. 1

- [52] Yi Yu, Xue Yang, Qingyun Li, Yue Zhou, Feipeng Da, and Junchi Yan. H2rbox-v2: Incorporating symmetry for boosting horizontal box supervised oriented object detection. In *Advances in Neural Information Processing Systems*, 2023. [1](#), [2](#), [3](#), [4](#), [6](#), [7](#)
- [53] Yi Yu, Xue Yang, Qingyun Li, Feipeng Da, Jifeng Dai, Yu Qiao, and Junchi Yan. Point2rbox: Combine knowledge from synthetic visual patterns for end-to-end oriented object detection with single point supervision. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16783–16793, 2024. [1](#), [3](#), [7](#)
- [54] Yi Yu, Botao Ren, Peiyuan Zhang, Mingxin Liu, Junwei Luo, Shaofeng Zhang, Feipeng Da, Junchi Yan, and Xue Yang. Point2rbox-v2: Rethinking point-supervised oriented object detection with spatial layout among instances. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025. [2](#), [3](#), [4](#), [6](#), [7](#)
- [55] Yi Yu, Xue Yang, Yansheng Li, Zhenjun Han, Feipeng Da, and Junchi Yan. Wholly-wood: Wholly leveraging diversified-quality labels for weakly-supervised oriented object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–18, 2025. [1](#), [2](#)
- [56] Dingwen Zhang, Junwei Han, Gong Cheng, and Ming-Hsuan Yang. Weakly supervised object localization and detection: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5866–5885, 2021. [2](#)
- [57] Peiyuan Zhang, Junwei Luo, Xue Yang, Yi Yu, Qingyun Li, Yue Zhou, Xiaosong Jia, Xudong Lu, Jingdong Chen, Xiang Li, et al. Pointobb-v3: Expanding performance boundaries of single point-supervised oriented object detection. *International Journal of Computer Vision*, 133:6108–6128, 2025. [7](#)
- [58] Shun Zhang, Jihui Long, Yaohui Xu, and Shaohui Mei. Pmho: Point-supervised oriented object detection based on segmentation-driven proposal generation. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. [3](#)
- [59] Like Zhao, Shunyi Zheng, Wenjing Yang, Haitao Wei, and Xia Huang. An image thresholding approach based on gaussian mixture model. *Pattern Analysis and Applications*, 22: 75–88, 2019. [5](#)
- [60] Yuanyi Zhong, Jianfeng Wang, Jian Peng, and Lei Zhang. Boosting weakly supervised object detection with progressive knowledge transfer. In *European Conference on Computer Vision*, pages 615–631. Springer, 2020. [2](#), [5](#)
- [61] Yue Zhou, Xue Yang, Gefan Zhang, Jiabao Wang, Yanyi Liu, Liping Hou, Xue Jiang, Xingzhao Liu, Junchi Yan, Chengqi Lyu, et al. Mmrotate: A rotated object detection benchmark using pytorch. In *30th ACM International Conference on Multimedia*, pages 7331–7334, 2022. [5](#)
- [62] Tianyu Zhu, Bryce Ferenczi, Pulak Purkait, Tom Drummond, Hamid Rezaatofghi, and Anton Van Den Hengel. Knowledge combination to learn rotated detection without rotated annotation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15518–15527, 2023. [2](#)