IMBALANCE-ADAPTIVE CROSS CONSISTENCY LEARNING FOR SEMI-SUPERVISED ORIENTED OBJECT DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Oriented object detection (OOD) has achieved significant progress under fully supervised learning. However, its performance heavily relies on large-scale annotated datasets, which are costly and time-consuming to obtain. To alleviate this limitation, semi-supervised oriented object detection (SSOOD) has attracted increasing attention due to its ability to leverage unlabeled data. Nevertheless, existing SSOOD methods often suffer from severe class imbalance in pseudo-label generation, particularly under long-tailed data distributions, which hinders performance. To address this issue, we propose Imbalance-Adaptive Cross Consistency Learning (IACCL), a novel SSOOD framework that simultaneously mitigates class imbalance and enhances model robustness. Specifically, we introduce a Balanced Pseudo-Label (BPL) strategy that adaptively regulates the number of pseudo-labels across categories based on their composite scores, thereby mitigating the dominance of head classes and promoting a more balanced learning process. In addition, we design a Cross Consistency Constraint (CCC) to enforce consistency between the teacher and student models under shared feature representations, encouraging the extraction of more robust and discriminative features, thereby alleviating overfitting under limited supervision. Extensive experiments on benchmark datasets demonstrate that the proposed method achieves consistent improvements over the state-of-the-art methods, particularly for tail categories, validating its effectiveness in handling class imbalance in SSOOD.

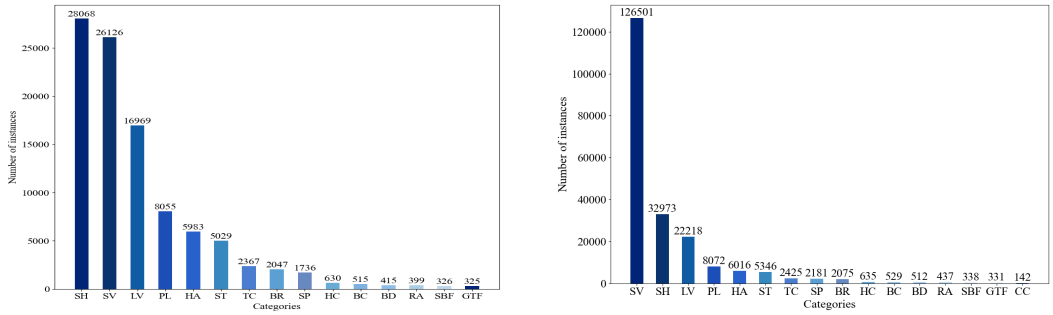
1 INTRODUCTION

Oriented object detection (OOD) has significant practical value and is widely applied in remote sensing imagery. Most existing research in OOD focuses on supervised learning, which relies heavily on large amounts of labeled data. However, in real-world scenarios, obtaining sufficient labeled data is often challenging. Consequently, recent studies have shifted toward semi-supervised learning (SSL), which aims to improve model performance by effectively leveraging unlabeled data.

Semi-supervised learning has achieved remarkable success in horizontal object detection (HOD). Representative methods such as Unbiased Teacher Liu et al. (2021), SoftTeacher Xu et al. (2021), Dense Teacher Zhou et al. (2022a), ARSL Liu et al. (2023), PseCo Li et al. (2022), and TMR-RD Marvasti-Zadeh et al. (2024) have demonstrated that unlabeled data can effectively narrow the performance gap between semi-supervised and fully supervised detectors. Existing semi-supervised oriented object detection (SSOOD) methods inherit the paradigms of HOD, with the primary adaptation involving the substitution of horizontal bounding boxes (HBBs) with oriented bounding boxes (OBBs) to handle the orientation-aware detection task. Typically, SSOOD approaches adopt a teacher-student paradigm: the student network learns from both labeled data and teacher-generated pseudo-labels, while the teacher network is updated via an exponential moving average (EMA) of the student's weights to ensure stability. To further enhance sample diversity, the teacher and student networks often process unlabeled images with weak and strong data augmentations, respectively, thereby facilitating the learning of robust, invariant representations from limited labeled data.

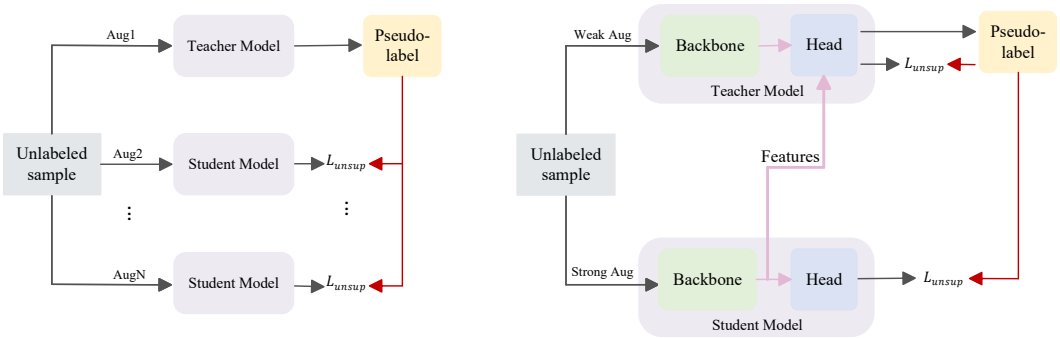
First, pseudo-label selection is heavily biased due to the severe class imbalance inherent in OOD datasets, which is frequently overlooked by current SSOOD approaches. Most methods rely on

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) Training-set distribution of DOTA v1.0. (b) Training-set distribution of DOTA v1.5.

Figure 1: Long-tailed category distributions in the DOTA v1.0 and DOTA v1.5 training sets.



(a) Multi-data Augmentation (b) Cross Consistency Constraint (Ours)

Figure 2: Comparison of consistency-learning paradigms. (a) Multi-view consistency uses additional augmentations to improve model robustness. (b) CCC applies shared pseudo-label supervision to the teacher and student heads on the same feature maps.

dense pseudo-labeling based solely on classification confidence Zhou et al. (2022a). However, as illustrated in Fig. 1, our statistical analysis of mainstream OOD datasets (DOTAv1.0 Xia et al. (2018) and DOTAv1.5 Ding et al. (2022)) reveals a significant long-tailed distribution: a small number of "head classes" account for the majority of instances, while "tail classes" contain extremely limited samples. Under such imbalanced conditions, the teacher network tends to generate pseudo-labels biased toward head classes. This bias further propagated during student network training, exacerbating the imbalance and ultimately leading to degraded performance on tail classes. To address this issue, we propose Balanced Pseudo-Label (BPL), which evaluates samples using a composite classification-regression score and adaptively suppresses pseudo-labels from dominant classes to achieve a more balanced distribution.

Second, the scarcity of labeled data in semi-supervised learning often limits the generalization ability of the model. A common approach to address this issue is to use different augmentations Berthelot et al. (2020); Li et al. (2021); Nguyen (2024) for the same image to generate more training images, as shown in Fig. 2(a). Unlike this approach, we propose the Cross Consistency Constraint (CCC) which focuses on improving the robustness of the feature representation. The feature maps generated by the backbone of the student model are fed into the detection heads of both the teacher and student models, as shown in Fig. 2(b). And the predictions of both models are supervised by the pseudo-label, which enforces the backbone to produce more representative features of objects. Our contributions can be summarized as follows.

- To address long-tailed distributions, we design a Balanced Pseudo-Label (BPL) strategy that applies an adaptive suppression factor to dominant classes, thereby balancing pseudo-label distributions across classes and alleviating the long-tailed effect.
- We propose the Cross Consistency Constraint to enforce pseudo-label-aligned consistency between teacher and student predictions, promoting more robust backbone representations.
- Our IACCL method achieves state-of-the-art performance in various experimental settings on the DOTA v1.0 and DOTA v1.5 benchmarks.

2 Related Work

Semi-Supervised Oriented Object Detection. Oriented object detection handles arbitrary object orientations, posing challenges in bounding-box representation and regression. Representative methods address these challenges through oriented proposal generation (Xie et al., 2021), hierarchical feature modeling (Yao et al., 2024), geometry-aware sample selection (Yao et al., 2025), and direction-aware regression (Song et al., 2024). However, they rely on extensive oriented-box annotations. Semi-supervised oriented object detection (SSOOD) alleviates this burden through teacher–student learning. Existing methods improve pseudo-label supervision through rotation-aware weighting and global consistency (Hua et al., 2023), dual teachers and probabilistic regression (Wu et al., 2024), focal learning (Wang et al., 2024a), dense multi-scale pseudo-labels (Zhao et al., 2025a), spatial consistency (Xia et al., 2025), contrastive background learning and label correction (Zhao et al., 2025b), resolution-aware supervision (Li et al., 2026), and joint representation and pseudo-label learning (Meng et al., 2026). However, class-wise imbalance in pseudo-label allocation remains underexplored. IACCL addresses this gap through BPL, which regulates category-wise pseudo-label budgets based on candidate prevalence and prediction quality, and CCC, which enforces cross-head consistency on shared feature maps.

Long-Tailed Training. Long-tailed object detection often favors head classes and provides insufficient supervision for tail classes. Existing approaches mainly use data- or optimization-level rebalancing. Forest R-CNN combines hierarchical category modeling with NMS resampling to increase the contribution of tail categories during training (Wu et al., 2020); Equalization Loss suppresses dominant negative gradients that hinder rare-category learning (Tan et al., 2020), whereas Equalization Loss v2 further balances category-wise positive and negative gradients based on accumulated statistics (Tan et al., 2021). In remote-sensing semi-supervised detection, Wang et al. select informative samples and decouple backbone and detector training to reduce category imbalance (Wang et al., 2023), while MOI extends this direction to oriented objects through iterative labeled-data filtering, phased training, and detection post-processing (Wang et al., 2024b). In SSOOD, imbalance can also arise during pseudo-label generation: frequent or well-learned categories tend to produce more confident predictions and dominate dense pseudo-label supervision. Whereas these approaches rebalance labeled samples or detector optimization, they do not directly regulate category-wise pseudo-label allocation during self-training. This distinction matters because the teacher's evolving predictions determine the class composition of unlabeled supervision and can reinforce existing skew; the resulting candidate distribution is dynamic rather than fixed by labeled-set frequencies. BPL addresses this source directly by estimating class-wise dominance from candidate prevalence and prediction quality, then adapting pseudo-label budgets without predefined head–tail partitions.

Consistency Regularization. Semi-supervised classification methods improve stability by enforcing consistency across augmented views. ReMixMatch uses Augmentation Anchoring to align strongly augmented views with a weakly augmented prediction (Berthelot et al., 2020); CoMatch models sample relationships through contrastive graph learning (Li et al., 2021); and SequenceMatch imposes consistency across weak, intermediate, and strong augmentations (Nguyen, 2024). In semi-supervised horizontal object detection, PseCo extends multi-view consistency to differently scaled inputs under pseudo-label supervision, improving robustness to object-scale changes (Li et al., 2022). These methods enforce consistency across views, augmentation strengths, or input scales. In contrast, CCC applies shared pseudo-label targets to the student and EMA-teacher heads using student features. This cross-head constraint encourages representations compatible with detection heads at different optimization states without extra inference computation.

3 Method

3.1 Overall Framework and Training Objective

Fig. 3 provides an overview of the proposed Imbalance-Adaptive Cross Consistency Learning (IACCL) method. It consists of a student network and a teacher network sharing the same architecture. During the semi-supervised training stage, the parameters of the student network are updated via backpropagation, while the teacher network parameters are updated as an exponential moving average (EMA) of the student parameters. Unlabeled data with weak augmentation are first fed into the teacher model, whose dense predictions are used to generate pseudo-labels via the Balanced

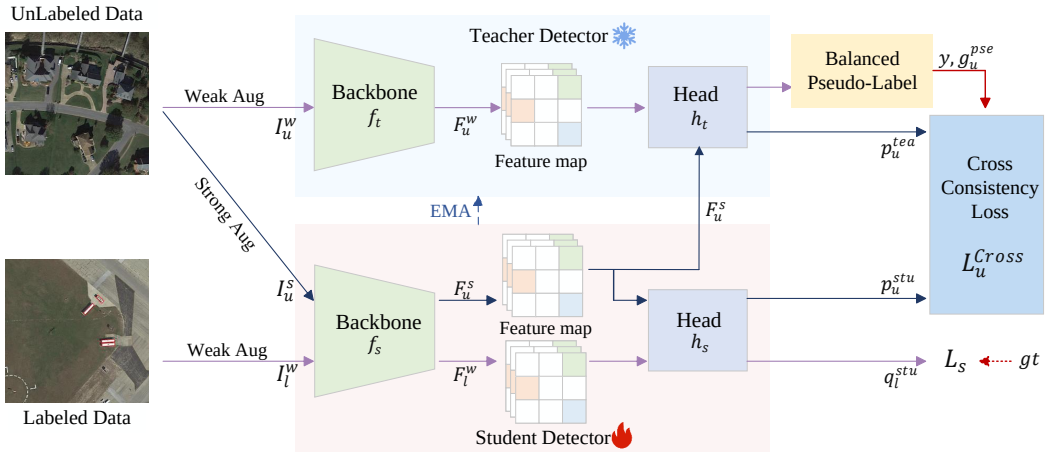


Figure 3: An illustration of our proposed IACCL

Pseudo-Label (BPL) strategy. This strategy alleviates the inherent bias toward head classes, thereby promoting a more balanced learning process. For the same unlabeled data, strong augmentation is applied before feeding them into the student model. After feature extraction by the feature pyramid network, the resulting features are forwarded to the detection heads of both the teacher and student models. By enforcing consistency between the predictions of different detection heads on the same feature maps, the backbone is encouraged to learn more representative features.

Let $D_L = \{(X_L^i, Y_L^i)\}_{i=1}^{N_L}$ and $D_U = \{X_U^j\}_{j=1}^{N_U}$ denote the labeled and unlabeled datasets, respectively, where N_L and N_U are the corresponding numbers of samples. In IACCL, two models with the same configuration serve as the teacher model and the student model, respectively. The overall training process is divided into two stages: a burn-in stage and a semi-supervised stage. During the burn-in stage, the oriented object detector is trained using labeled data to obtain an initial student model, after which its parameters are copied to initialize the teacher model. In the semi-supervised stage, the student model updates its parameters using both labeled data and pseudo-labels generated by the teacher model. The overall loss for the student model throughout the training process is defined as:

$$L(t) = L_s + \lambda_u L_u^{\text{cross}}, \tag{1}$$

where L_s and L_u^{cross} are the supervised loss of labeled data and the unsupervised loss of unlabeled data, respectively. λ_u is a weight factor used to balance the contributions of different loss terms, and is set to 1.0 in this work. To enhance model robustness and reduce the negative impact of inaccurate pseudo-labels, we follow previous work Hua et al. (2023) and update the parameters of the teacher model using EMA of the student model, as defined in Eq.(2).

$$\theta_{t+1}^{tea} = \alpha\theta_t^{tea} + (1-\alpha)\theta_t^{stu} \tag{2}$$

where θ_t^{tea} and θ_t^{stu} represent the parameters of the teacher model and the student model in the t-th iteration. α is set to 0.9996, following the setting in SOOD Hua et al. (2023) .

3.2 Balanced Pseudo-Label

Dense pseudo-labeling methods typically retain a fixed proportion of teacher predictions according to classification confidence (Zhou et al., 2022a; Zhao et al., 2025a). Under a long-tailed class distribution, frequent or well-learned categories tend to produce more high-confidence predictions and can therefore dominate pseudo-label supervision. Moreover, classification confidence alone does not adequately reflect localization quality. To address these limitations, we propose Balanced Pseudo-Label (BPL), which jointly evaluates the semantic and localization quality of each candidate and dynamically allocates class-wise pseudo-label budgets according to the current prediction distribution.

Given N dense candidate predictions produced by the teacher from a weakly augmented view of an unlabeled image, let $\mathbf{c}_k = [c_{k,1}, \ldots, c_{k,M}]$ denote the classification confidence vector of candidate k, where M is the number of classes, and let r_k denote its predicted rotated IoU quality. We assign

candidate k to the class with the highest classification confidence:

$$\hat{y}_k = \arg \max_{i \in \{1,\dots,M\}} c_{k,i}. \tag{3}$$

Its candidate composite score is then defined as

$$o_k = c_{k,\hat{y}_k} r_k. \tag{4}$$

The product favors candidates that are reliable in both category prediction and localization, rather than ranking them by classification confidence alone.

We group candidates according to their predicted classes: $J_i = \{k \mid \hat{y}_k = i\}$, and denote the set of classes present in the current candidate pool by $\mathcal{C} = \{i \mid |J_i| > 0\}$. For each $i \in \mathcal{C}$, we define its candidate prevalence ρ_i, mean candidate quality q_i, and class-wise dominance score s_i as

$$\rho_i = \frac{|J_i|}{N}, \qquad q_i = \frac{1}{|J_i|} \sum_{k \in J_i} o_k, \qquad s_i = \rho_i q_i = \frac{1}{N} \sum_{k \in J_i} o_k. \tag{5}$$

Here, ρ_i measures how frequently class i occurs in the candidate pool, whereas q_i measures its average prediction quality. Their product s_i quantifies the aggregate confidence mass contributed by class i. A large s_i therefore indicates that the class may contribute a disproportionate amount of pseudo-label supervision.

BPL determines dominant classes online from the current predictions, without requiring a predefined head–tail partition. Specifically, we compute the mean dominance score over the classes present in the candidate pool:

$$\bar{s} = \frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} s_i, \tag{6}$$

and assign each class an adaptive suppression factor:

$$w_i = \begin{cases} 1, & s_i \le \bar{s}, \\ \sqrt{\bar{s}/s_i}, & s_i > \bar{s}. \end{cases} \tag{7}$$

Thus, classes whose dominance scores do not exceed the current mean remain unsuppressed, whereas classes contributing excessive confidence mass receive $w_i < 1$.

The pseudo-label budget for class i is

$$d_i = \min \left\{ |J_i|, \lfloor \beta w_i s_i N \rfloor \right\}, \tag{8}$$

where β controls the overall retention scale and the outer minimum ensures that the budget does not exceed the number of available candidates. We select the d_i candidates with the highest o_k from J_i as pseudo-labels. In this way, BPL preserves candidates with jointly high semantic confidence and localization quality while reducing the relative contribution of categories that dominate the current prediction pool. The complete procedure is given in Appendix A.

3.3 Cross Consistency Constraint

In conventional teacher–student detectors, pseudo-label supervision for a strongly augmented unlabeled image is applied only through the student detection head. Although the EMA teacher head generates pseudo-targets from the weak view, the head itself is not used to process the student features. We therefore propose the Cross Consistency Constraint (CCC), which reuses the frozen teacher head as an auxiliary supervision path. CCC imposes cross-head supervision on the student features without introducing an additional learnable head or changing the inference architecture.

Given an unlabeled image I_u, the teacher processes its weakly augmented view I_u^w to generate dense predictions. After BPL selection, the pseudo-targets $g_u^{\text{pse}} = (g_{\text{cls}}^{\text{pse}}, g_{\text{reg}}^{\text{pse}}, g_{\text{iou}}^{\text{pse}})$ are aligned with the strongly augmented view I_u^s. The student feature extractor, including the backbone and FPN, produces

$$F_u^s = f_s(I_u^s). \tag{9}$$

CCC feeds this feature map to both the student head h_s and the frozen teacher head h_t:

$$\begin{aligned} p_u^{\text{stu}} &= (p_{\text{cls}}^{\text{stu}}, p_{\text{reg}}^{\text{stu}}, p_{\text{iou}}^{\text{stu}}) = h_s(F_u^s), \\ \tilde{p}_u^{\text{tea}} &= (\tilde{p}_{\text{cls}}^{\text{tea}}, \tilde{p}_{\text{reg}}^{\text{tea}}, \tilde{p}_{\text{iou}}^{\text{tea}}) = h_t(F_u^s). \end{aligned} \tag{10}$$

Table 1: Comparison under the Partially Labeled Data setting on DOTA v1.0 and v1.5. Bold and underlined values denote the best and second-best semi-supervised results, respectively. $*$, †, and ‡ denote Rotated Faster R-CNN, Rotated FCOS, and Oriented R-CNN, respectively; "–" denotes an unreported result.

		DOTA v1.5			DOTA v1.0		
Setting	Method	10%	20%	30%	10%	20%	30%
Supervised	Faster R-CNN$*$ (Ren et al., 2017)	43.43	51.32	53.14	44.96	55.98	57.89
	FCOS† (Tian et al., 2019)	42.78	50.11	54.79	44.83	54.16	57.41
Semi-supervised	Unbiased Teacher$*$ (Liu et al., 2021)	44.51	52.80	53.33	–	–	–
	Soft Teacher$*$ (Xu et al., 2021)	48.46	54.89	57.83	–	–	–
	Dense Teacher† (Zhou et al., 2022a)	46.90	53.93	57.86	50.87	58.02	60.91
	PseCo$*$ (Li et al., 2022)	48.04	55.28	58.03	–	–	–
	DualPolish$*$ (Zhang et al., 2023)	49.02	55.17	58.44	–	–	–
	SOOD† (Hua et al., 2023)	48.63	55.58	59.23	52.11	59.23	61.44
	ARSL† (Liu et al., 2023)	47.56	55.05	58.94	52.89	59.84	62.92
	PST$*$ (Wu et al., 2024)	49.63	57.39	60.40	<u>55.79</u>	<u>62.51</u>	<u>65.94</u>
	MOI† (Wang et al., 2024b)	50.57	57.13	61.04	–	–	–
	Focal Teacher† (Wang et al., 2024a)	52.24	57.00	60.21	–	–	–
	TMR-RD-v2$*$ (Marvasti-Zadeh et al., 2024)	47.88	55.17	59.02	52.81	59.90	63.12
	Denser Teacher† (Zhao et al., 2025a)	52.05	57.49	60.40	52.62	59.20	62.82
	SRODET† (Wang et al., 2025)	49.23	55.38	59.16	–	–	–
	Adjacent Teacher† (Xia et al., 2025)	51.26	57.42	60.26	–	–	–
	SAOD† (Zhu et al., 2025)	51.01	57.14	60.44	–	–	–
	DCPLS† (Yang et al., 2025)	49.23	55.71	59.10	–	–	–
	CBLC-SOOD† (Zhao et al., 2025b)	49.64	56.12	59.36	52.77	58.70	62.79
	Oriented Teacher‡ (Li et al., 2026)	<u>52.30</u>	<u>57.76</u>	<u>61.23</u>	–	–	–
	SROD† (Meng et al., 2026)	50.54	57.60	60.95	52.58	55.12	59.41
	IACCL (ours)†	**54.00**	**60.05**	**61.99**	**58.08**	**63.68**	**66.77**

The notation $\tilde{p}_u^{\mathrm{tea}}$ distinguishes the teacher-head prediction on the student feature map from the weak-view teacher prediction used to construct the pseudo-targets.

For a head prediction $p = (p_{\mathrm{cls}}, p_{\mathrm{reg}}, p_{\mathrm{iou}})$, we define

$$\ell_u(p, g_u^{\mathrm{pse}}) = L^{\mathrm{QFL}}(p_{\mathrm{cls}}, g_{\mathrm{cls}}^{\mathrm{pse}}) + L^{\mathrm{RIoU}}(p_{\mathrm{reg}}, g_{\mathrm{reg}}^{\mathrm{pse}}) + L^{\mathrm{BCE}}(p_{\mathrm{iou}}, g_{\mathrm{iou}}^{\mathrm{pse}}). \tag{11}$$

where L^{QFL}, L^{RIoU}, and L^{BCE} denote the Quality Focal Loss (QFL) (Li et al., 2020), rotated IoU loss, and binary cross-entropy loss, respectively. Following Dense Teacher (Zhou et al., 2022a), the BPL-selected locations use the teacher classification probabilities as the soft QFL targets, whereas unselected locations are assigned zero targets. The regression and quality losses are evaluated only at the selected locations, with the regression loss additionally weighted by the teacher-predicted quality scores.

Using the same BPL-selected targets for both heads, CCC is formulated as

$$L_u^{\mathrm{cross}} = \ell_u(p_u^{\mathrm{stu}}, g_u^{\mathrm{pse}}) + \ell_u(\tilde{p}_u^{\mathrm{tea}}, g_u^{\mathrm{pse}}). \tag{12}$$

The first term is the conventional student-head loss, whereas the second introduces the proposed cross-head supervision. Because h_t is frozen, gradients from the second term pass through the teacher head and update only the student feature extractor. The student features are consequently optimized to support predictions under both the current student head and the EMA-stabilized teacher head. This shared-target supervision induces cross-head consistency without an explicit pairwise discrepancy loss between their predictions. At inference, only the student detector is retained, so CCC adds no inference parameters or computation. The resulting L_u^{cross} is used in the overall objective in Eq. equation 1.

4 EXPERIMENTS

4.1 DATASETS AND EVALUATION PROTOCOLS

We conduct experiments on DOTA v1.0 and DOTA v1.5, two standard benchmarks for evaluating semi-supervised oriented object detection (SSOOD). **DOTA v1.0** (Xia et al., 2018) contains 2,806 aerial images, divided into 1,411 training, 458 validation, and 937 test images, with 188,282 oriented bounding-box annotations across 15 categories. **DOTA v1.5** (Ding et al., 2022) retains the same image split, adds annotations for objects smaller than 10 pixels, and introduces container crane (CC) as a 16th category, yielding 403,318 annotated instances.

Following SOOD (Hua et al., 2023), we consider the Partially Labeled Data and Fully Labeled Data protocols. In the **Partially Labeled Data setting**, 10%, 20%, or 30% of the training images are used as labeled data, while the remaining training images are treated as unlabeled. We use fixed nested splits for these labeling ratios: the 10% labeled set is contained in the 20% set, which is contained in the 30% set. This protocol provides a common data basis for controlled comparisons across annotation budgets. In the **Fully Labeled Data setting**, all training images are labeled, whereas the test images are used only as unlabeled inputs; their annotations are never accessed during training.

Under both settings, we report mean average precision (mAP) on the corresponding validation set, following the standard DOTA evaluation protocol and the practice of prior SSOOD studies.

4.2 IMPLEMENTATION DETAILS

IACCL is implemented in PyTorch with MMRotate (Zhou et al., 2022b). Rotated FCOS (Tian et al., 2019) with an FPN serves as the base detector, together with an ImageNet-pretrained ResNet-50 backbone (Deng et al., 2009). Following SOOD (Hua et al., 2023), each image is cropped into 1024×1024 patches with a stride of 824, corresponding to an overlap of 200 pixels. Training runs for 120k iterations on a single NVIDIA GeForce RTX 3090 with a batch size of 3 (two labeled and one unlabeled image). Stochastic gradient descent (SGD) is used with an initial learning rate of 6.25×10^{-4}, a weight decay of 1×10^{-4}, and a momentum of 0.9. The burn-in period comprises 6,400 iterations, and the unsupervised-loss weight is set to $\lambda_u = 1.0$.

For unlabeled images, strong augmentation is applied to the student input and weak augmentation to the teacher input. The strong augmentation pipeline includes random flipping, random rotation, random grayscale conversion, color jitter, and random Gaussian blur, whereas the weak augmentation pipeline includes only random flipping and random rotation.

4.3 COMPARISON WITH THE STATE-OF-THE-ART METHODS

Comparative Results on Partially Labeled Data. Table 1 compares IACCL with recent SSOOD methods on DOTA v1.5 and DOTA v1.0. On DOTA v1.5, IACCL achieves 54.00, 60.05, and 61.99 mAP with 10%, 20%, and 30% labeled data, respectively, improving over the supervised Rotated FCOS baseline by 11.22, 9.94, and 7.20 mAP and over the second-best methods by 1.70, 2.29, and 0.76 mAP. On DOTA v1.0, it reaches 58.08, 63.68, and 66.77 mAP, with corresponding gains of 13.25, 9.52, and 9.36 mAP over the supervised baseline and margins of 2.29, 1.17, and 0.83 mAP over the second-best methods. IACCL therefore ranks first among the compared methods at every labeling ratio on both datasets. Qualitative comparisons with Dense Teacher (Zhou et al., 2022a) and SOOD (Hua et al., 2023) under the 30% labeled setting on DOTA v1.5 are provided in Appendix D (Figure 7). The examples illustrate differences in missed detections, false positives, and localization quality among the compared methods.

Comparative Results on Fully Labeled Data. Table 2 reports semi-supervised results alongside their supervised baselines under the Fully Labeled Data setting. On DOTA v1.5, IACCL improves the Rotated FCOS baseline from 65.46 to 67.95 mAP, ranking second and falling 0.10 mAP below Oriented Teacher, which uses a different base detector. On DOTA v1.0, it raises the baseline from 70.22 to 72.32 mAP and achieves the best result, surpassing TMR-RD-v2 by 0.64 mAP.

Additional qualitative results under the Fully Labeled Data setting on DOTA v1.0 are provided in Appendix D (Figure 8). Additional evaluations of transfer to horizontal object detection, cross-detector generalization, and inference efficiency are provided in Appendix C.

Table 2: Results under the Fully Labeled Data setting on DOTA v1.0 and v1.5. Bold and underlined values denote the best and second-best semi-supervised results, respectively. Symbols follow Table 1.

Method	DOTA v1.5	DOTA v1.0
Unbiased Teacher∗ (Liu et al., 2021)	66.12 $\xrightarrow{-1.27}$ 64.85	–
Soft Teacher∗ (Xu et al., 2021)	66.12 $\xrightarrow{+0.28}$ 66.40	–
Dense Teacher† (Zhou et al., 2022a)	65.46 $\xrightarrow{+0.92}$ 66.38	70.22 $\xrightarrow{+0.98}$ 71.20
SOOD† (Hua et al., 2023)	65.46 $\xrightarrow{+2.24}$ 67.70	–
ARSL† (Liu et al., 2023)	65.46 $\xrightarrow{+1.22}$ 66.68	70.22 $\xrightarrow{+1.41}$ 71.63
Focal Teacher† (Wang et al., 2024a)	65.46 $\xrightarrow{+1.66}$ 67.12	–
TMR-RD-v2∗ (Marvasti-Zadeh et al., 2024)	66.12 $\xrightarrow{+0.81}$ 66.93	70.08 $\xrightarrow{+1.60}$ <u>71.68</u>
SRODET† (Wang et al., 2025)	65.34 $\xrightarrow{+1.78}$ 67.12	–
Adjacent Teacher† (Xia et al., 2025)	66.23 $\xrightarrow{+1.16}$ 67.39	–
AODet† (Chen et al., 2025)	65.46 $\xrightarrow{+2.32}$ 67.78	–
SAOD† (Zhu et al., 2025)	65.46 $\xrightarrow{+2.36}$ 67.82	–
CBLC-SOOD† (Zhao et al., 2025b)	65.46 $\xrightarrow{+1.15}$ 66.61	70.22 $\xrightarrow{+0.45}$ 70.67
Oriented Teacher‡ (Li et al., 2026)	66.12 $\xrightarrow{+1.93}$ **68.05**	–
IACCL (ours)†	65.46 $\xrightarrow{+2.49}$ <u>67.95</u>	70.22 $\xrightarrow{+2.10}$ **72.32**

Table 3: Ablation of BPL and CCC.

BPL	CCC	DOTA v1.0	DOTA v1.5
–	–	52.88	50.19
✓	–	57.15	53.25
–	✓	54.42	51.30
✓	✓	**58.08**	**54.00**

Table 4: Ablation of BPL score terms.

Cls.	IoU	mAP
✓	–	52.71
–	✓	50.97
✓	✓	**53.25**

Table 5: Sensitivity to β.

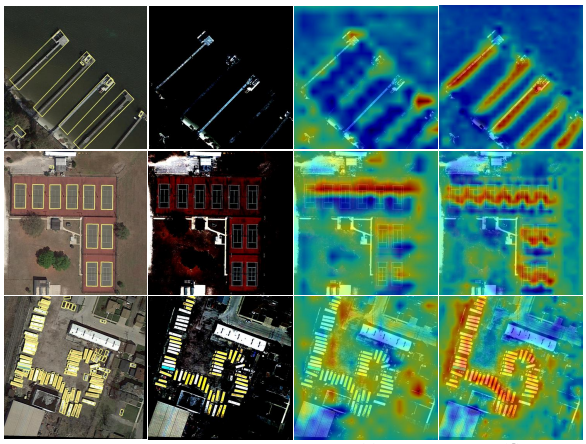
β	mAP
8	52.45
10	52.88
12	52.56
14	53.10
16	**53.25**
18	52.54

4.4 Ablation Study

To isolate component contributions, ablations are conducted under the 10% labeled setting on both DOTA v1.0 and DOTA v1.5; the remaining experiments use the 10% labeled setting on DOTA v1.5.

Component Effectiveness. Table 3 reports 52.88 and 50.19 mAP for the baseline without BPL and CCC on DOTA v1.0 and DOTA v1.5, respectively. Adding BPL increases mAP by 4.27 and 3.06 on DOTA v1.0 and DOTA v1.5, respectively, whereas CCC alone increases mAP by 1.54 and 1.11 on the two datasets. Combining BPL and CCC reaches 58.08 and 54.00 mAP, providing additional gains of 0.93 and 0.75 over BPL alone and total gains of 5.20 and 3.81 over the baseline, respectively. These results show that BPL and CCC make complementary contributions. The corresponding qualitative comparison of FPN features with and without CCC is shown in Figure 4.

BPL Design Choices. Table 4 evaluates the candidate composite score used by BPL. Using classification confidence or IoU quality alone yields 52.71 and 50.97 mAP, respectively, whereas their product reaches 53.25 mAP, improving over the two single-term variants by 0.54 and 2.28 mAP. This result indicates that pseudo-label selection benefits from jointly considering semantic confidence and localization quality. Table 5 examines the effect of the retention scale β. Performance peaks at $\beta = 16$ with 53.25 mAP; smaller values retain fewer candidates, whereas larger values can admit more lower-quality predictions, resulting in lower mAP. Appendix B provides further analysis of the mechanism. Figure 6 shows that suppression progressively reduces pseudo-labels for head and medium categories while increasing the allocation for most tail categories, with the redistribution becoming more pronounced at the later training checkpoint. Figure 5 reports the corresponding category-wise AP changes: most tail categories improve, while head and medium categories vary



(a) (b) (c) (d)

Figure 4: Qualitative comparison of FPN features with and without CCC. (a) Original images with ground-truth annotations. (b) Strongly augmented images. (c) FPN features learned without CCC. (d) FPN features learned with CCC.

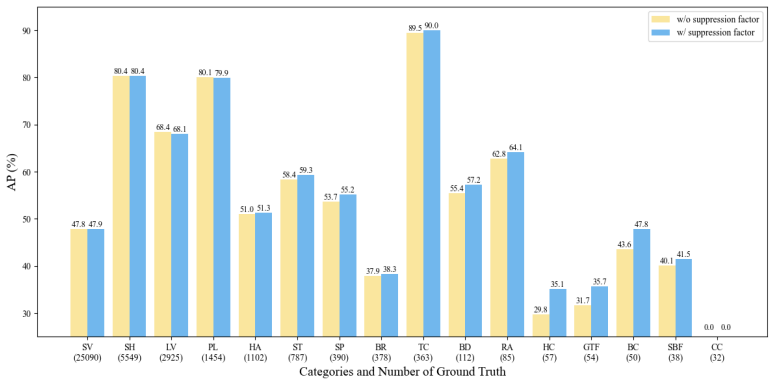


Figure 5: Category-wise AP with and without the suppression factor on DOTA v1.5 (10% labeled).

only slightly. Together, these results support the intended role of BPL in reducing the relative contribution of categories with high dominance scores while directing pseudo-label supervision toward underrepresented categories.

Category-wise AP Analysis. To evaluate the effect of pseudo-label redistribution, we compare category-wise AP with and without the suppression factor under the 10% labeled setting on DOTA v1.5. As shown in Fig. 5, most tail categories improve, with gains of up to 5.3 points for HC and 4.2 points for BC. In contrast, head and medium categories change only marginally, while CC remains at 0 AP due to extremely scarce labeled instances. These results demonstrate that the suppression factor effectively alleviates head-class dominance and improves learning for underrepresented categories.

5 Conclusion

We presented IACCL, a semi-supervised oriented object detection framework designed to address class imbalance and limited supervision in SSOOD. BPL estimates class-wise dominance from candidate prevalence and prediction quality, and adaptively regulates category-wise pseudo-label selection to reduce the influence of dominant categories. CCC applies shared pseudo-label supervision to the teacher and student detection heads on the same student features, encouraging cross-head consistency and more robust feature learning. Across DOTA v1.0 and DOTA v1.5, IACCL achieves the best reported results under all evaluated partially labeled settings and consistently improves the supervised baseline under the Fully Labeled Data setting. These results support the effectiveness of combining class-adaptive pseudo-label selection with cross-head consistency supervision.

AI USE STATEMENT

In preparing this manuscript, the authors used generative AI tools for language editing and polishing. The authors did not use generative AI tools to generate research ideas, formulate or prove mathematical claims, design or conduct experiments, generate or clean datasets, implement the proposed method, or interpret experimental results. All AI-assisted outputs were reviewed and verified by the authors, including the manuscript text, citations, technical claims, equations, and experimental descriptions. The authors take full responsibility for the final content of this work.

REFERENCES

David Berthelot, Nicholas Carlini, Ekin D. Cubuk, Alex Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring. In *International Conference on Learning Representations*, 2020.

Jiecong Chen, Chenlin Fu, and Yingying Zhu. Adaptive optimization strategy for semi-supervised arbitrary-oriented object detection. In *2025 IEEE International Conference on Multimedia and Expo*, pp. 1–6, 2025. doi: 10.1109/ICME59968.2025.11210067.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

Jian Ding, Nan Xue, Gui-Song Xia, Xiang Bai, Wen Yang, Michael Ying Yang, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Object detection in aerial images: A large-scale benchmark and challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7778–7796, 2022. doi: 10.1109/TPAMI.2021.3117983.

W. Hua, D. Liang, J. Li, X. Liu, Z. Zou, X. Ye, and X. Bai. Sood: Towards semi-supervised oriented object detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15558–15567, 2023. ISBN 2575-7075. doi: 10.1109/CVPR52729.2023.01493.

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 Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (eds.), *Computer Vision – ECCV 2022*, pp. 457–472. Springer Nature Switzerland, 2022. ISBN 978-3-031-20077-9.

Junnan Li, Caiming Xiong, and Steven C.H. Hoi. Comatch: Semi-supervised learning with contrastive graph regularization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9475–9484, 2021.

Xiang Li, Wenhai Wang, Lijun Wu, Shuo Chen, Xiaolin Hu, Jun Li, Jinhui Tang, and Jian Yang. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, volume 33, pp. 21002–21012, 2020.

Yuqiu Li, Jian Xue, and Ke Lu. Oriented teacher: Exploiting spatial resolution disparity to enhance semi-supervised aerial object detection. *IEEE Open Journal of Signal Processing*, 7:472–481, 2026. doi: 10.1109/OJSP.2026.3685499.

C. Liu, W. Zhang, X. Lin, W. Zhang, X. Tan, J. Han, X. Li, E. Ding, and J. Wang. Ambiguity-resistant semi-supervised learning for dense object detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15579–15588, 2023. ISBN 2575-7075. doi: 10.1109/CVPR52729.2023.01495.

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. In *International Conference on Learning Representations*, 2021.

Seyed Mojtaba Marvasti-Zadeh, Nilanjan Ray, and Nadir Erbilgin. Training-based model refinement and representation disagreement for semi-supervised object detection. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2245–2254, 2024.

Lingqiang Meng, Shengkai Gao, Xin Lyu, Hongmin Gao, Zhizhong Liu, Yiwei Fang, Zhennan Xu, and Xin Li. Jointly enhancing representation and supervision for semisupervised rotated object detection of remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 19:25850–25870, 2026. doi: 10.1109/JSTARS.2026.3718516.

K. B. Nguyen. Sequencematch revisiting the design of weak-strong augmentations for semi-supervised learning. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 96–105, 2024. ISBN 2642-9381. doi: 10.1109/WACV57701.2024.00017.

S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (6):1137–1149, 2017. ISSN 1939-3539. doi: 10.1109/TPAMI.2016.2577031.

Beihang Song, Jing Li, Jia Wu, Jun Chang, and Jun Wan. Direction prediction redefinition: Transfer angle to scale in oriented object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(12):12894–12906, 2024. doi: 10.1109/TCSVT.2024.3438431.

J. Tan, C. Wang, B. Li, Q. Li, W. Ouyang, C. Yin, and J. Yan. Equalization loss for long-tailed object recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11659–11668, 2020. ISBN 2575-7075. doi: 10.1109/CVPR42600.2020.01168.

J. Tan, X. Lu, G. Zhang, C. Yin, and Q. Li. Equalization loss v2: A new gradient balance approach for long-tailed object detection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1685–1694, 2021. ISBN 2575-7075. doi: 10.1109/CVPR46437.2021.00173.

Z. Tian, C. Shen, H. Chen, and T. He. Fcos: Fully convolutional one-stage object detection. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9626–9635, 2019. ISBN 2380-7504. doi: 10.1109/ICCV.2019.00972.

K. Wang, Z. Xiao, Q. Wan, F. Xia, P. Chen, and D. Li. Global focal learning for semi-supervised oriented object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024a. ISSN 1558-0644. doi: 10.1109/TGRS.2024.3438844.

Wenyong Wang, Yuanzheng Cai, and Tao Wang. SRODET: Semi-supervised remote sensing object detection with dynamic pseudo-labeling. *IEEE Geoscience and Remote Sensing Letters*, 22:1–5, 2025. doi: 10.1109/LGRS.2025.3544807.

Y. Wang, L. Yao, G. Meng, X. Zhang, J. Song, and H. Zhang. Semi-supervised object detection in remote sensing images based on active learning. In *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5571–5574, 2023. ISBN 2153-7003. doi: 10.1109/IGARSS52108.2023.10282261.

Y. Wang, L. Yao, X. Zhang, J. Song, and H. Zhang. Mining oriented information for semi-supervised object detection in remote sensing images. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2024b. ISBN 2161-4407. doi: 10.1109/IJCNN60899.2024.10650864.

Jialian Wu, Liangchen Song, Tiancai Wang, Qian Zhang, and Junsong Yuan. Forest r-cnn: Large-vocabulary long-tailed object detection and instance segmentation. In *Proceedings of the 28th ACM international conference on multimedia*, pp. 1570–1578, 2020.

W. Wu, H. S. Wong, and S. Wu. Pseudo-siamese teacher for semi-supervised oriented object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024. ISSN 1558-0644. doi: 10.1109/TGRS.2024.3380645.

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 *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3974–3983, 2018. doi: 10.1109/CVPR.2018.00418.

Tao Xia, Wei Jing, and Qi Wang. Adjacent teacher: Semi-supervised oriented object detection leveraging adjacent spatial consistency prior in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–13, 2025. doi: 10.1109/TGRS.2025.3579430.

X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han. Oriented r-cnn for object detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3500–3509, 2021. ISBN 2380-7504. doi: 10.1109/ICCV48922.2021.00350.

M. Xu, Z. Zhang, H. Hu, J. Wang, L. Wang, F. Wei, X. Bai, and Z. Liu. End-to-end semi-supervised object detection with soft teacher. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3040–3049, 2021. ISBN 2380-7504. doi: 10.1109/ICCV48922.2021.00305.

Zewei Yang, Jie Feng, Junpeng Zhang, Weisheng Dong, and Ronghua Shang. Dual-criterion pseudo label selection for semi-supervised oriented object detection on remote sensing images. In *2025 IEEE International Geoscience and Remote Sensing Symposium*, pp. 8506–8510, 2025. doi: 10.1109/IGARSS55030.2025.11243394.

Yanqing Yao, Gong Cheng, Chunbo Lang, Xiang Yuan, Xingxing Xie, and Junwei Han. Hierarchical mask prompting and robust integrated regression for oriented object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(12):13071–13084, 2024. doi: 10.1109/TCSVT.2024.3444795.

Yanqing Yao, Gong Cheng, Chunbo Lang, Xingxing Xie, and Junwei Han. Centric probability-based sample selection for oriented object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(4):3290–3302, 2025. doi: 10.1109/TCSVT.2024.3509872.

Lei Zhang, Yuxuan Sun, and Wei Wei. Mind the gap: Polishing pseudo labels for accurate semi-supervised object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 3463–3471, 2023.

Tong Zhao, Qiang Fang, and Xin Xu. Denser teacher: Rethinking dense pseudo-label for semi-supervised oriented object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(5):4549–4559, 2025a. doi: 10.1109/TCSVT.2024.3518452.

Ye Zhao, Yong Zhu, Yuxi Gong, Xueliang Liu, and Liangfeng Xu. CBLC-SOOD: Contrastive background and label correction for semi-supervised oriented object detection. *Multimedia Systems*, 31(6):438, 2025b. doi: 10.1007/s00530-025-02028-5.

Hongyu Zhou, Zheng Ge, Songtao Liu, Weixin Mao, Zeming Li, Haiyan Yu, and Jian Sun. Dense teacher: Dense pseudo-labels for semi-supervised object detection. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (eds.), *Computer Vision – ECCV 2022*, pp. 35–50. Springer Nature Switzerland, 2022a. ISBN 978-3-031-20077-9.

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 *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 7331–7334, 2022b.

Ye Zhu, Yan Bai, Shixin Cen, and Cuihong Xue. Towards semi-supervised adaptive oriented object detection. *Remote Sensing Letters*, 16(6):665–675, 2025. doi: 10.1080/2150704X.2025.2492903.

A Balanced Pseudo-Label Algorithm

Algorithm 1 gives the complete pseudo-label selection procedure used by BPL. It first combines classification confidence and localization quality at the candidate level, estimates class-wise dominance from candidate prevalence and mean quality, and then retains the highest-scoring candidates under an adaptive budget for each predicted class.

Algorithm 1 Balanced Pseudo-Label

Require: Number of classes M; number of teacher predictions N; predicted rotated IoU quality scores $\{r_k\}_{k=1}^{N}$; classification score vectors $\{\mathbf{c}_k\}_{k=1}^{N}$; scaling factor β
Ensure: Predicted classes $\{\hat{y}_k\}_{k=1}^{N}$ and selection masks $\{m_k\}_{k=1}^{N}$

for $i = 1$ to M **do**
  $J_i \leftarrow \emptyset$
end for
for $k = 1$ to N **do**
  $\hat{y}_k \leftarrow \arg\max_{i \in \{1,\dots,M\}} c_{k,i}$
  $o_k \leftarrow c_{k,\hat{y}_k} \cdot r_k$
  $J_{\hat{y}_k} \leftarrow J_{\hat{y}_k} \cup \{k\}$; $m_k \leftarrow 0$
end for
$\mathcal{C} \leftarrow \{i \mid |J_i| > 0\}$
for $i \in \mathcal{C}$ **do**
  $\rho_i \leftarrow |J_i|/N$
  $q_i \leftarrow |J_i|^{-1} \sum_{k \in J_i} o_k$
  $s_i \leftarrow \rho_i q_i$
end for
$\bar{s} \leftarrow |\mathcal{C}|^{-1} \sum_{i \in \mathcal{C}} s_i$
for $i \in \mathcal{C}$ **do**
  $w_i \leftarrow 1$
  if $s_i > \bar{s}$ **then**
    $w_i \leftarrow \sqrt{\bar{s}/s_i}$
  end if
  $d_i \leftarrow \min\{|J_i|, \lfloor \beta w_i s_i N \rfloor\}$
  Sort the candidates in J_i by descending composite score o_k
  for $j = 1$ to d_i **do**
    Set the mask of the j-th ranked candidate in J_i to 1
  end for
end for

B Additional Analysis of BPL: Pseudo-Label Redistribution

We analyze how the suppression factor changes the class distribution of pseudo-labels under the 10% labeled setting on DOTA v1.5. Models are trained with and without the suppression factor, and their pseudo-label distributions are evaluated on the validation set. For each category, we compute the pseudo-label enhancement ratio (PLER), defined as the number of pseudo-labels generated with suppression divided by the number generated without suppression. A PLER above 1 indicates an increase in pseudo-labels, whereas a value below 1 indicates suppression.

Figure 6 compares PLER at 57,600 and 102,400 training iterations. As training progresses, suppression becomes stronger for head and medium categories, substantially reducing their pseudo-label counts. In contrast, most tail categories receive more pseudo-labels, and this redistribution becomes more pronounced at the later checkpoint. BD and RA remain comparatively suppressed despite receiving additional pseudo-labels because they retain relatively high dominance scores in the current candidate pool. Overall, the observed distribution shift supports the intended mechanism of BPL: limiting categories with high aggregate confidence mass and directing training toward underrepresented categories.

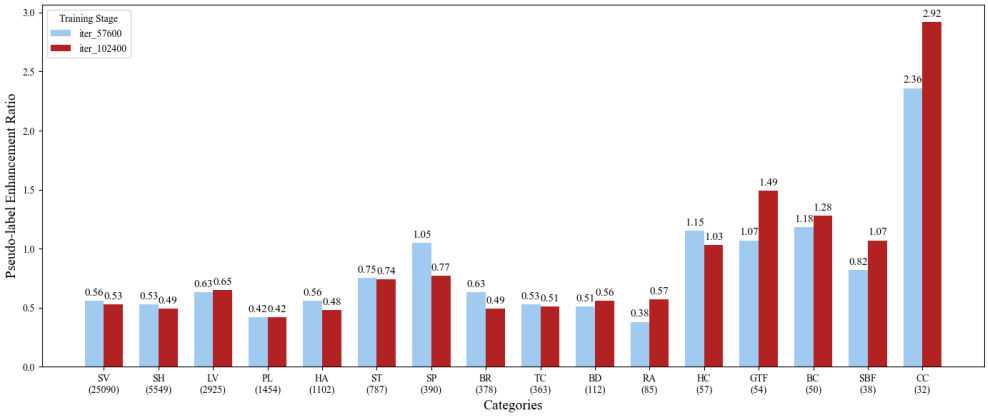


Figure 6: Pseudo-label enhancement ratios induced by the suppression factor at 57,600 and 102,400 training iterations under the 10% labeled setting on DOTA v1.5. A value above 1 indicates an increase in pseudo-labels, whereas a value below 1 indicates suppression.

C Additional Generalization and Efficiency Results

C.1 Transfer to Horizontal Object Detection

Table 6: Results on DOTA v1.5 using horizontal bounding boxes. Full denotes the Fully Labeled Data setting; Δ denotes reproduced results.

Setting	Method	10%	20%	30%	Full
Supervised	FCOS† Δ (Tian et al., 2019)	45.8	55.2	58.8	69.7
Semi-supervised	Dense Teacher† Δ (Zhou et al., 2022a)	47.8	58.3	60.8	67.5
	IACCL (ours)†	**54.2**	**62.0**	**64.3**	**71.2**

Table 6 evaluates whether the proposed training strategy transfers from oriented to horizontal object detection. IACCL outperforms Dense Teacher by 6.4, 3.7, 3.5, and 3.7 mAP under the 10%, 20%, 30%, and Fully Labeled Data settings, respectively. The consistent gains indicate that the method is not tied to oriented-box parameterization.

C.2 Generalization Across Detectors

Table 7: Cross-detector generalization on DOTA v1.5 under the 10% labeled setting. Results are mAP.

Method	Rotated RetinaNet	Rotated FCOS
Baseline	48.76	50.19
Ours	49.72	54.00

As shown in Table 7, applying IACCL to Rotated RetinaNet improves mAP from 48.76 to 49.72 (0.96 mAP), while applying it to Rotated FCOS improves mAP from 50.19 to 54.00 (3.81 mAP). Improvements with both detectors support the architectural generality of the training framework, although the gain depends on the underlying detector.

C.3 Computational Efficiency

Table 8 shows that SOOD†, Dense Teacher†, and IACCL† have identical inference complexity because they use the same detector architecture. IACCL† therefore raises mAP without increasing

Table 8: Detection accuracy and inference efficiency on DOTA v1.5 under the 10% labeled setting.

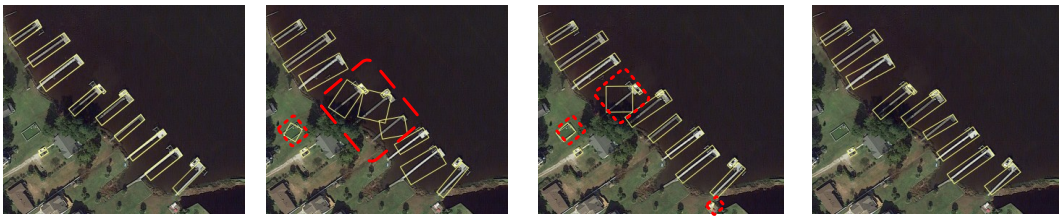
Method	mAP	GFLOPs	Params (M)	FPS
SOOD† (Hua et al., 2023)	48.63	206.91	31.92	17.7
Dense Teacher† (Zhou et al., 2022a)	46.90	206.91	31.92	17.7
IACCL†	54.00	206.91	31.92	17.7
IACCL°	49.72	215.92	36.42	18.7

GFLOPs, parameter count, or inference latency relative to these FCOS-based methods. The Rotated RetinaNet variant, IACCL°, uses more parameters and GFLOPs but reaches a slightly higher measured FPS. These measurements characterize inference efficiency; the teacher–student overhead incurred during training is not included.

D Additional Qualitative Results

This appendix presents the qualitative detection results referenced in the main paper.

D.1 Partially Labeled Data



(a) Ground Truth (b) Dense Teacher (c) SOOD (d) Ours

Figure 7: Visualization results under the 30% labeled setting on DOTA v1.5. Red annotations indicate inaccurate predictions.

D.2 Fully Labeled Data



Figure 8: Qualitative detection results under the Fully Labeled Data setting on DOTA v1.0.