ACTIVE-SAOOD: ACTIVE SPARSELY ANNOTATED ORIENTED OBJECT DETECTION IN REMOTE SENSING IMAGES

Anonymous authors
Paper under double-blind review

ABSTRACT

Sparse annotation has emerged as an effective strategy for reducing the annotation burden of oriented object detection in remote sensing images. However, with a limited annotation budget, selecting informative instances becomes critical, while random selection may allocate annotations to less valuable instances. To this end, we propose an active learning-based sparsely annotated oriented object detection (SAOOD) method, termed Active-SAOOD, which actively selects high-value instances for annotation. Specifically, existing image-level active learning methods face two limitations in SAOOD task: (1) the granularity mismatch in annotation value evaluation and (2) unstable training caused by excessive selection of difficult samples without considering model states. To overcome these challenges, we design instance-level evaluation dimensions tailored to SAOOD, enabling the selection of instances that improve both fundamental detection and representation generalization abilities. Moreover, a model-state observation strategy is introduced to identify instances that better match the current learning state of model, thereby improving training stability. Extensive experiments on multiple datasets demonstrate that the Active-SAOOD consistently improves detection performance and training stability under different sparse annotation ratios, highlighting its practical value for SAOOD. The code will be public.

1 INTRODUCTION

Oriented object detection (OOD) in remote sensing images is a task with significant practical importance Han et al. (2021a); Yang et al. (2021), but its application is hindered by the expensive manual annotation. To reduce the labeling cost, recent works have explored weakly supervised Yang et al. (2022a), semi-supervised Hua et al. (2023), and sparsely annotated Liao et al. (2025) approaches. Sparsely annotated oriented object detection (SAOOD) has attracted increasing attention for well reducing the annotation redundancy in dense scenes.

Although recent SAOOD methods have significantly reduced annotation costs by sparsely labeling instances, they ignore that annotated instances have different learning values. Under limited annotation budgets, randomly selected samples may waste annotation resources on low-value instances while missing those that can substantially improve detection ability. Therefore, identifying valuable instances is crucial for maintaining SAOOD performance under low annotation rates.

Existing active learning methods are mainly explored in image classification Li et al. (2024) and semi-supervised object detection Mi et al. (2022), where annotation value is typically modeled at the image-level. However, this paradigm is not well suited for SAOOD: (1) SAOOD assigns annotations to object instances rather than images, requiring fine-grained instance-level annotation value estimation. We refer to this setting as instance-level active learning. (2) Remote sensing images suffer from class imbalance and large intra-class variations, making instance value should consider not only on classification difficulty but also on category coverage and intra-class representation. Moreover, the additional angle prediction in oriented object detection introduces orientation-related annotation value that remains unexplored. (3) Image-level active learning can select the most difficult images since each image contains diverse instances, naturally ensuring diverse instance difficulties. In contrast, instance-level active learning cannot simply select the hardest instances, as heavily

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.

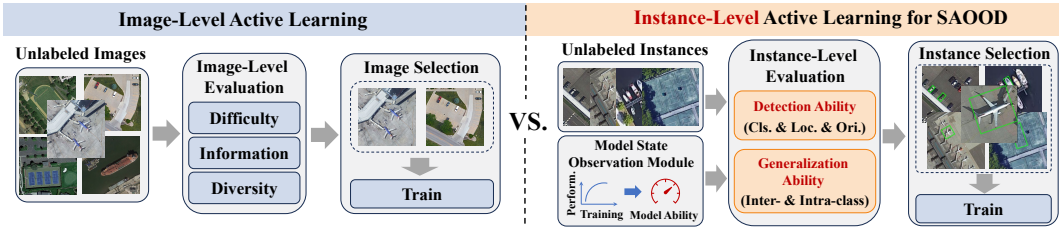


Figure 1: Comparison of sampling methods. Image-level active learning selects valuable images based on image-level evaluation, while instance-level active learning evaluates finer-grained instance value and selects instances according to model state. Active-SAOOD outperforms random annotation by selecting high-value instances.

sampling difficult ones may destabilize early training, which is widely demonstrated in curriculum learning Wang et al. (2021). Therefore, SAOOD requires a task-specific instance value modeling strategy beyond conventional image-level sample selection.

To address the above challenges, we propose Active-SAOOD, a fine-grained instance-level active learning framework for SAOOD. First, to bridge the gap between image-level active learning and instance-level annotation, we revisit SAOOD from the perspective of instance annotation value and decompose it into three fundamental detection ability dimensions: classification, localization, and orientation prediction. Based on this decomposition, we define corresponding instance-level uncertainty measures to select samples that improve the detector along different capability aspects. Second, to better model the representational characteristics of remote sensing data, we propose representation generalization ability, which evaluates how instances enrich the annotated data representation space. Specifically, inter-class diversity is used to improve category coverage and alleviate long-tailed distributions, while intra-class diversity selects feature-diverse instances to enhance generalization ability and address large intra-class variations in remote sensing. Third, to avoid the instability caused by directly selecting the hardest instances, we design a model state observation strategy that dynamically adjusts the importance of different evaluation dimensions according to the current learning status of the detector, enabling a progressive learning process from easy to hard instances. Finally, these components jointly form a unified instance-level active learning framework, enabling Active-SAOOD to achieve improved performance and training stability under extremely sparse annotations. Our main contributions are as follows:

1) We propose Active-SAOOD, a novel instance-level active learning framework for sparsely annotated oriented object detection. We revisit SAOOD from the perspective of model capabilities and decompose it into two components: fundamental detection ability and representation generalization ability. Based on the characteristics of SAOOD and remote sensing images, we propose corresponding instance-level evaluation dimensions, addressing the limitation that conventional image-level active learning is not well suited for SAOOD.
2) We design a model state observation strategy that adaptively balances fundamental detection and representation generalization ability learning across different training stages, selecting high-value instances suited to the current model state.
3) Extensive experiments demonstrate that Active-SAOOD improves SAOOD performance and training stability under extremely sparse annotations, further enhancing the practical value of SAOOD.

2 Related Work

Oriented Object Detection. Oriented object detection (OOD) predicts rotated boxes to explicitly model object orientations and is widely used in remote sensing He et al. (2017); Xie et al. (2021). Early methods Ma et al. (2018) extended horizontal detectors, while later works improved remote sensing detection through better feature representation (R^3Det Yang et al. (2021)), anchor design (S^2A-Net Han et al. (2021a)), and proposal generation (Oriented R-CNN Xie et al. (2021)). Other studies focus on boundary discontinuity Yang et al. (2022b); Yu & Da (2023), rotation-equivariant

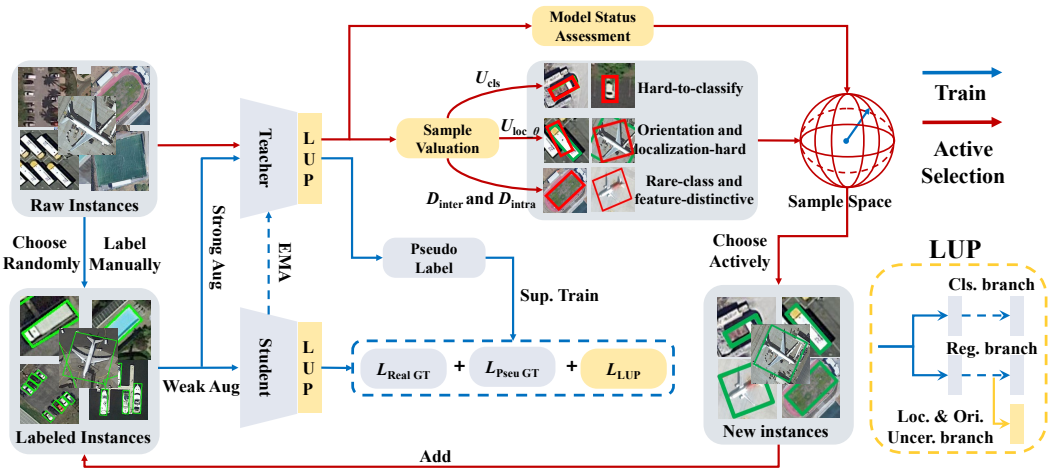


Figure 2: Overall framework of Active-SAOOD. A small randomly sparsely annotated set initializes the teacher-student models. The model state observation strategy evaluates the fundamental detection and representation generalization abilities of teacher, guiding the selection of high-value instances that best fit the current learning state, progressively enhancing the training set.

representation Han et al. (2021b), and receptive field enhancement Li et al. (2025). Despite these advances, OOD remains annotation-intensive, driving low-cost annotation research.

Sparsely Annotated Oriented Object Detection. Recently, sparsely annotated oriented object detection (SAOOD) has gained attention for reducing annotation redundancy in dense remote sensing scenes. PECL Lu et al. (2024) used conformal learning for high-quality pseudo-labels, DML Zhang et al. (2025) applied dynamic multi-view learning, S^2Teacher Lin et al. (2026) leveraged angle-consistency and progressive mining, and RSST Liao et al. (2025) utilized large language models for pseudo-label assignment. However, these methods rely on strong category-dependent priors, making random sparse annotations unstable, and the value of sparse samples is largely unexplored. To address this, we propose Active-SAOOD, which actively selects high-value samples for sparse annotation, ensuring more stable performance of SAOOD.

Active Learning. Active learning reduces annotation costs by selecting the most informative samples. Existing studies in image classification Wang et al. (2016) and semi-supervised object detection focus on image-level selection, with Active Teacher Mi et al. (2022) selecting valuable images, MI-AOD Yuan et al. (2021) estimating uncertainty through adversarial classifier discrepancy, ALWOD Wang et al. (2023) using teacher–student disagreement, and PPAL Yang et al. (2024) providing a plug-and-play module. However, sparse annotation requires evaluating instance-level value, and existing methods overlook orientation uncertainty, which is crucial for OOD. We introduce multi-dimensional instance-level annotation value, including orientation, to improve SAOOD.

3 ACTIVE-SAOOD

Given a small initial sparsely annotated instance set $\mathcal{D}_I = \{\mathcal{X}_I, \mathcal{Y}_I\}$ and an unlabeled candidate instance set $\mathcal{D}_U = \{\mathcal{X}_U\}$, where $\mathcal{X}$ denotes instances and $\mathcal{Y}$ denotes labels, Active-SAOOD aims to select the valuable instances from $\mathcal{D}_U$ for annotation and then merge them with $\mathcal{D}_I$ to form an actively annotated set $\mathcal{D}_A$, which is used to train the detector for improving performance.

As shown in Figure 2, Active-SAOOD initializes a teacher–student SAOOD framework with a small set of randomly sparse annotations. Teacher model evaluates the annotation value of samples based on proposed instance-level evaluation dimension. Guided by model state observation strategy, it selects instances most beneficial to current model state. These selected instances are manually annotated and merged with the initial training-set to iteratively train student model. Active-SAOOD targets high-value instance selection rather than exhaustive object discovery, with iterative candidate regeneration enabling previously missed objects to become selectable as the teacher improves.

Algorithm 1 Pseudo Code of Active-SAOOD

Require: Sparse annotated set $\mathcal{D}_I$; Unlabeled instance set $\mathcal{D}_U$; Maximum iteration N
Ensure: Trained detector M^t
1: Train initial student model M_0^s on $\mathcal{D}_I$ and initialize teacher model M_0^t via EMA
2: **for** $n = 1, \ldots, N$ **do**
3: Observe model state $\mathcal{O}_{n-1}$ via MSO from M_{n-1}^t.
4: Compute annotation value $S(x_u|\mathcal{O}_{n-1})$ for $x_u \in \mathcal{D}_U^{n-1}$ using M_{n-1}^t via Eq. (11)
5: Select top-K instances $\mathcal{X}_A^k$ with highest scores and annotate labels $\mathcal{Y}_A^k$
6: Update labeled set $\mathcal{D}_A^n = \mathcal{D}_A^{n-1} \cup \{\mathcal{X}_A^k, \mathcal{Y}_A^k\}$
7: Train student model M_n^s on $\mathcal{D}_A^n$ and update teacher model M_n^t via EMA
8: **end for**
9: **return** M_N^t

3.1 Instance-Level Active Sampling for SAOOD

Existing active learning methods mainly focus on image classification and semi-supervised object detection, and cannot be directly applied to SAOOD due to two limitations. First, they evaluate annotation value at the image level, while SAOOD requires fine-grained instance-level evaluation. Second, they do not consider the angular dimension in oriented object detection and its impact on annotation value, as well as the characteristics of remote sensing scenarios.

To address these limitations, we revisit what makes an instance valuable for annotation under sparse annotation. We argue that **instance annotation value should reflect its contribution to improving the ability of current detector**. Therefore, we decompose SAOOD into fundamental detection abilities, including classification, localization, and orientation, and propose corresponding instance-level uncertainty measures to guide abilities enhancement. Furthermore, we propose inter-class and intra-class diversity to adapt to remote sensing characteristics, considering long-tailed distributions and large intra-class variations while selecting instances that improve rare category recognition and intra-class robustness. These components jointly form an instance-level active learning framework.

3.2 Fundamental Detection Ability Evaluation

Based on the objectives of SAOOD task, we decompose fundamental detection ability evaluation into: instance-level classification, localization, and orientation uncertainty.

Classification Uncertainty: Uncertainty is a widely used measure in active learning. Following prior work Wu et al. (2022), we adopt the information entropy of the predicted class probability distribution from the teacher model as the instance-level classification uncertainty. Specifically, for each object, the teacher model outputs a normalized probability distribution $p = [p_1, p_2, \ldots, p_C]$, where p_c represents the probability that the object belongs to class c. The instance-level classification uncertainty is defined as the information entropy:

$$U_{\text{cls}} = -\sum_{c=1}^{C} p_c \log(p_c), \tag{1}$$

Higher entropy indicates greater class prediction uncertainty and lower classification confidence.

Localization and Orientation Uncertainty: In oriented object detection, localization involves both bounding box regression and orientation prediction. Existing active learning methods mainly focus on image-level annotation value, with limited exploration of instance-level localization and orientation uncertainty. Therefore, we propose an instance-level localization and orientation uncertainty. Since regression branches lack probabilistic distribution outputs, it is difficult to directly assess the localization accuracy of predicted boxes. Therefore, we design a plug-and-play localization and orientation uncertainty prediction (LUP) module with minimal training overhead. As shown in Fig. 2, LUP adds a parallel uncertainty branch to the regression head, consisting of a single 3×3 convolution with one output channel to predict uncertainty for each rotated bounding box.

During training, the LUP requires supervision. We use Rotated Intersection over Union (RIoU) between predicted and ground-truth boxes as localization uncertainty supervision:

$$U_{\text{loc}} = 1 - \text{RIoU}(B_{\text{pred}}, B_{\text{gt}}), \tag{2}$$

where B_{pred} and B_{gt} denote the predicted and ground-truth boxes, respectively. A larger RIoU corresponds to lower localization uncertainty, indicating that the object is easier to localize and provides limited contribution to localization ability improving. Conversely, highly uncertain instances are more valuable for enhancing the localization ability.

Although RIoU implicitly reflects orientation accuracy, it is a coupled measure of localization and orientation errors. Previous studies Zeng et al. (2024a) show that even for elongated objects with a 5:1 aspect ratio, RIoU can remain above 0.5 under angular deviations below 15°, indicating that RIoU alone cannot accurately measure angular errors. Therefore, we further propose orientation uncertainty by measuring the angular deviation between predicted and ground-truth boxes. However, oriented object detection suffers from the edge-swapping problem. When the predicted and ground-truth boxes have different long-side orientations, they can be geometrically equivalent under a 90° rotation. Thus, we first align the long-side directions of predicted and ground-truth boxes before computing angular deviation. For a predicted box $(x_p, y_p, w_p, h_p, \theta_p)$ and a ground-truth $(x_g, y_g, w_g, h_g, \theta_g)$, where x, y denote center coordinates, w, h denote width and height, and θ denotes angle, we first determine whether their long-side directions are consistent:

$$s_p = [w_p \geq h_p], \quad s_g = [w_g \geq h_g], \quad \theta'_p = \begin{cases} \theta_p + \frac{\pi}{2}, & s_p \oplus s_g = 1, \\ \theta_p, & \text{otherwise.} \end{cases} \tag{3}$$

Here, $[\cdot]$ denotes the indicator function and $\oplus$ denotes the XOR operation. When $s_p \oplus s_g = 1$, the two boxes are geometrically equivalent under a 90° rotation, so we adjust the angle accordingly.

After correction, the predicted box is aligned with the ground truth in long-side orientation, and then the angles are normalized to $[-\frac{\pi}{2}, \frac{\pi}{2})$. Due to the angle boundary problem in oriented object detection Xu et al. (2024), directly computing angular differences may introduce inconsistencies between numerical and actual deviations, leading to unstable uncertainty prediction. Therefore, we use the minimal angular difference as the supervision signal for angle uncertainty:

$$U_\theta = \min\left(|\theta'_p - \theta_g|, \pi - |\theta'_p - \theta_g|\right), \tag{4}$$

where θ'_p is the corrected predicted angle and θ_g is the ground-truth angle. Low orientation uncertainty indicates that the predicted orientation closely matches ground truth, representing an easy-to-learn instance for orientation. Conversely, high-uncertainty instances are more challenging and provide greater potential for improving orientation ability.

Since objects with larger aspect ratios are more sensitive to angular deviations Zeng et al. (2024b), orientation uncertainty should be assigned higher weights for such objects. Accordingly, the final instance-level localization and orientation uncertainty is defined as:

$$U_{\text{loc_}\theta} = w\, U_{\text{loc}} + (1 - w)\, \frac{U_\theta}{\pi/2}, \quad w = e^{-\beta\left|\log\left(\frac{h_o}{w_o}\right)\right|}, \tag{5}$$

where w_o and h_o denote the object width and height, respectively, and β is the aspect ratio sensitivity factor, set to 0.5 in this paper. This formulation adaptively increases the orientation uncertainty weight for objects with larger aspect ratios and emphasizes localization uncertainty for near-square objects. See supplementary material for detailed analysis.

Since the localization and orientation uncertainty is normalized to $[0, 1]$ and treated as soft probability targets, cross-entropy loss enables more stable optimization:

$$\mathcal{L}_{\text{LUP}} = -\sum \Big[U_{\text{loc_}\theta} \log \hat{U}_{\text{loc_}\theta} + (1 - U_{\text{loc_}\theta}) \log(1 - \hat{U}_{\text{loc_}\theta})\Big], \tag{6}$$

where $\hat{U}_{\text{loc_}\theta}$ denotes the predicted localization and orientation uncertainty for each instance.

3.3 Representation Generalization Ability Evaluation

Based on the characteristics of remote sensing scenarios, we decompose representation generalization ability evaluation into inter-class and intra-class diversity.

Inter-Class Diversity: Remote sensing images often exhibit classes inbalance, which hinder rare category detection accuracy, especially under low sparse annotation rates where annotated rare instances are extremely limited. To address this, we propose instance-level inter-class diversity to increase sampling priority of rare categories. Specifically, we define inter-class diversity using a global class-frequency adjustment factor. Let N_c denote labeled instances number for class c, and D_{inter} is defined as a sigmoid-like function:

$$D_{\text{inter}} = \frac{1}{1 + e^{(\gamma N_c - 1)}}, \tag{7}$$

where γ is the smoothing coefficient, which is set to 0.01 in this work. Rare classes with fewer instances obtain higher D_{inter}, encouraging selection of these classes to improve the category diversity, while well-represented classes receive lower values to avoid redundant sampling.

Intra-Class Diversity: To handle large intra-class variations in remote sensing, we propose intra-class diversity to encourage rich and representative feature coverage within class. The mean classification feature vector of initially annotated instances is used as the class prototype. During active learning, cosine similarity between an instance and its prototype measures intra-class diversity. We define D_{intra} and dynamically update the class prototype of selected instances as follows:

$$D_{\text{intra}} = 1 - \frac{\mathbf{f}_i \cdot \mathbf{P}_c}{\|\mathbf{f}_i\| \, \|\mathbf{P}_c\|}, \quad \mathbf{P}_c \leftarrow \alpha \mathbf{P}_c + (1 - \alpha)\mathbf{f}_i. \tag{8}$$

where $\mathbf{f}_i$ is the classification feature of the i-th instance and $\mathbf{P}_c$ is the prototype feature of its class. A higher D_{intra} indicates that the instance contributes novel information to the feature space, helping the model learn more diverse features and improve generalization. Here, α is the momentum, set to 0.9 in this work. This strategy smooths prototype updates and ensures that prototypes gradually reflect the evolving feature distribution of annotated instances.

3.4 Model State Observation Strategy

For image-level active learning, directly selecting the most difficult images is sufficient, as each image contains diverse instances with varying difficulty. However, in instance-level active learning, selecting too many difficult instances in early training stages may harm training stability Wang et al. (2021). Curriculum learning demonstrates that gradually learning from easy to hard samples is more effective. Moreover, since the importance of each evaluation dimension varies across training stages, simply summing them may be suboptimal for the current model. To address these issues, we propose a model state observation strategy (MSO).

The core idea of MSO is that annotation value should depend not only on instance difficulty, but more importantly on its potential contribution to improving the ability of current detector. Specifically, MSO evaluates the model state from fundamental detection and representation generalization abilities, guiding the selection of instances aligned with current learning state. Fundamental detection ability corresponds to the core objectives of oriented object detection, including classification and localization, while representation generalization ability reflects robust and diverse feature learning in remote sensing through inter- and intra-class diversity. Classification ability is measured by mean Average Precision (mAP), localization ability by the mean IoU of predictions matched with sparse GT, inter-class diversity by rare-class recognition ability, and intra-class diversity by intra-class robustness. The four-dimensional abilities are computed as follows:

$$\begin{aligned} A_{\text{cls}} &= \frac{1}{C}\sum_{c=1}^{C} AP_c, \quad A_{\text{loc}} = \frac{1}{N}\sum_{i=1}^{N} \text{IoU}\left(b_i^{\text{pred}}, b_i^{\text{GT}}\right), \\ A_{\text{inter}} &= \sum_{c \in \text{rare}} AP_c, \quad A_{\text{intra}} = \frac{1}{C}\sum_{c=1}^{C} \text{Acc}_c \cdot \exp(-\text{Var}_c), \end{aligned} \tag{9}$$

where C is the number of classes, N the number of matched predictions, AP_c the average precision of class c, b_i^{pred} and b_i^{GT} the predicted and GT boxes, Acc_c the class accuracy, and Var_c the confidence variance. **Notably**, these abilities are not designed to measure absolute detector performance, but to characterize **relative learning progress** across different dimensions, guiding sample selection.

Table 1: Comparison of our Active-SAOOD with other SAOOD and image-level active learning (AL) methods on DOTA-v1.0 under different annotation rate. mAP is used for evaluation, and Δ denotes the mAP gain over the baseline. Active-Rotated FCOS is trained only with actively selected high-value samples, while Active-SAOOD additionally incorporates mined pseudo-labels.

Method paradigm	Method	Sparse annotation rate							
		1%	Δ	2%	Δ	5%	Δ	10%	Δ
Baseline	Rotated FCOS	20.94	+0.00	33.23	+0.00	48.59	+0.00	55.06	+0.00
SAOOD	DML Zhang et al. (2025)	21.62	+0.68	34.91	+1.68	52.51	+3.92	59.16	+4.10
	RSST Liao et al. (2025)	23.19	+2.25	37.20	+3.97	55.22	+6.63	60.62	+5.56
	S^2Teacher Lin et al. (2026)	30.73	+9.79	37.86	+4.63	55.49	+6.90	61.67	+6.61
Image-level AL	Active Teacher Mi et al. (2022)	29.20	+8.26	37.55	+4.32	55.06	+6.47	60.72	+5.66
	PPAL Yang et al. (2024)	30.71	+9.77	37.62	+4.39	56.23	+7.64	60.75	+5.69
Instance-level AL	Active-Rotated FCOS	30.24	+9.30	35.42	+2.19	53.89	+5.30	57.90	+2.84
	Active-SAOOD (DML-based)	32.24	+11.30	40.49	+7.26	56.12	+7.53	60.12	+5.06
	Active-SAOOD (RSST-based)	**38.54**	+17.6	**43.09**	+9.86	57.99	+9.40	61.16	+6.10
	Active-SAOOD (S^2Teacher-based)	35.79	+14.85	42.98	+9.75	**58.39**	+9.80	**62.61**	+7.55

Before computing the composite score, all values are normalized to $[0, 1]$. We use softmax normalization over inverse model ability to derive dimension weights. This ensures that (1) all dimension weights are positive and normalized, and (2) dimensions with weaker current ability receive higher importance without manually designed thresholds. The integrated value is defined as follows:

$$S = \sum_i \frac{e^{1-A_i}}{\sum_j e^{1-A_j}} U_i, \tag{10}$$

where $i \in \{\text{cls}, \text{loc_}\theta, \text{inter}, \text{intra}\}$, and U_i denotes classification, localization and orientation uncertainty, as well as inter- and intra-class diversity. Eq. 10 enables sample selection to adapt to the current abilities across different dimensions. To further achieve progressively selecting more difficult instances as the model ability improves, we introduce an ability-aware weighting strategy. When the overall ability $\bar{A}$ is low, the contribution of $(1 - S)$ is increased, favoring the selection of relatively easier instances for stable training. As $\bar{A}$ increases, the contribution gradually shifts from $(1 - S)$ to S, resulting in an adaptive transition from easier to harder instances. This provides a curriculum-like sampling mechanism that progressively increases the difficulty of selected samples as detector becomes strong. The final sample score is defined as:

$$S_{\text{final}} = \bar{A} \cdot S + (1 - \bar{A}) \cdot (1 - S), \tag{11}$$

where $\bar{A}$ is the average model ability across all dimensions. Based on the ability values measured by MSO, weaker ability dimensions are assigned higher weights during sample selection, enabling the selected samples to jointly improve fundamental detection and representation generalization. Meanwhile, the sampling strategy progressively increases instance difficulty as the overall model ability improves, favoring relatively easier instances for stable optimization in early training and gradually shifting toward harder instances. This ability-aware curriculum improves training stability and performance. **More MSO analysis is provided in the supplementary material.**

4 Experiments

4.1 Dataset and Metric

We evaluate our method on three oriented object detection datasets: DOTA-v1.0 Xia et al. (2018), containing 2,806 images and 188,282 instances across 15 categories (cropped into 1024×1024 patches), DOTA-v1.5 Xia et al. (2018), annotating more small objects, and DIOR Cheng et al. (2022), containing 23,463 images and 192,472 instances across 20 categories. For SAOOD, we randomly select p% of instances without class priors to construct sparse annotations, with only the constraint that each category has at least one instance to avoid zero-shot detection. Following prior work Lu et al. (2024), we use mAP_{50} for evaluation.

Table 2: Generalization experiments on different datasets. SAOOD refers to S^{2}Teacher method.

Dataset	Method	Annotation Rate	
		1%	5%
DOTA-v1.5	Rotated FCOS	21.91	42.12
	SAOOD	27.61	47.72
	Active-SAOOD (S^{2}Teacher)	**33.72**	**50.76**
DIOR	Rotated FCOS	15.64	33.41
	SAOOD	23.17	37.06
	Active-SAOOD (S^{2}Teacher)	**25.58**	**40.73**

Table 3: Ablation study of evaluation dimensions on DOTA-v1.0 with 1% annotation rate.

U_{cls}	U_{loc_θ}	D_{inter}	D_{intra}	**MSO**	**mAP(%)**
✓	✓	✓	✓	✓	**35.79**
–	✓	✓	✓	✓	33.62
✓	–	✓	✓	✓	34.75
✓	✓	–	✓	✓	32.66
✓	✓	✓	–	✓	33.90
✓	✓	✓	✓	–	32.70

4.2 Experimental Settings

Following previous work Hua et al. (2023), we use Rotated FCOS Tian et al. (2019) as the baseline. Implementation and hyperparameters follow default mmrotate settings. For active learning, following the previous work Mi et al. (2022), algorithm 1 runs for 2 iterations. The annotation ratio p% denotes the total manually annotated instances, including initial and actively selected ones, and all methods use the same annotation budget. Experiments are conducted on four NVIDIA RTX 3090 GPUs with batch size 4. Following Xu et al. (2021), the EMA momentum is set to 0.9996.

4.3 Main Results

To validate the effectiveness of Active-SAOOD, we conduct experiments on DOTA-v1.0 under different sparse annotation rates. As shown in Table 1, compared with random annotation, Active-SAOOD evaluates instances from multiple dimensions and selects samples that best match the current model state through MSO. This ensures selected samples are valuable and aligned with model capability. At an extremely low annotation rate (1%), the baseline with random annotation achieves 20.94% mAP, while Active-SAOOD improves it to 30.24%. For SAOOD methods, Active-SAOOD also consistently improves performance: RSST with random annotation obtains 23.19% mAP, while Active-SAOOD (RSST-based) reaches 38.54%, and S^{2}Teacher improves from 30.73% to 35.79%. These results demonstrate that random annotation may select less informative instances, whereas Active-SAOOD better utilizes the annotation budget by prioritizing high-value samples, improving both performance and stability. Furthermore, when image-level active learning (e.g., Active Teacher and PPAL) are used for sample selection and S^{2}Teacher adopted for pseudo-label mining, the mAP under 1% annotation is only 29.2% and 30.71%, respectively, compared with 35.79% achieved by our instance-level Active-SAOOD. This verifies that conventional image-level evaluation and selection strategies are not suited for sparse annotation, resulting in selected samples that are less compatible with model state. These results highlight the advantage of selection guided by MSO.

4.4 Generalization Experiment on DIOR

To evaluate generalization, we conduct experiments on the DOTA-v1.5 and DIOR datasets. DOTA-v1.5 contains more small objects, while DIOR has more object categories but fewer instances, making valuable instance selection more challenging. As shown in Table 2, Active-SAOOD consistently improves SAOOD performance across different datasets and annotation rates, demonstrating its strong generalization. More generalization results are provided in the supplementary material.

4.5 Ablation Study

To verify the effectiveness of the proposed instance-level evaluation dimensions, we conduct ablation experiments on DOTA-v1.0 with 1% annotations. As shown in Table 3, removing any dimension degrades performance, as different dimensions improve corresponding model abilities. Classification, localization, and orientation uncertainty enhance fundamental detection ability, while inter- and intra-class diversity improve representation generalization by alleviating class imbalance and intra-class feature variations. MSO further improves selection by adapting to the current model state. **See supplementary material for more experiments**, including MSO evaluation on a held-out set, sample difficulty trends, hyperparameter studies, and more.



(a) Visualization of the actively selected instances.



(b) Comparison of detection results between models trained with randomly selected and actively selected instances.

Figure 3: Visualization analysis. (a) Red boxes are predicted boxes, while green boxes are ground-truth. (b) Green boxes are predicted boxes, and red circles highlight missed objects.

Table 4: Repeated experiments on DOTA-v1.0 with 1% annotation. The var denotes the variance.

Method	max	min	mean	var
Random annotation	20.94	4.42	13.45	71.16
SAOOD	30.73	12.15	26.25	62.59
Active-FCOS	30.24	25.52	27.35	4.29
Active-SAOOD	**35.79**	**30.46**	**32.80**	**4.08**

4.6 Repeat Annotation Experiment

To evaluate stability, we conduct five repeated experiments on DOTA with a 1% annotation rate. As shown in Table 4, random annotation exhibits large variations, with mAP ranging from 4.42% to 20.94% and a variance of 71.16, while SAOOD with pseudo-label mining remains unstable (variance is 62.59). In contrast, Active-SAOOD achieves stable performance, with mAP ranging from 30.46% to 35.79% and a variance of 4.08, improving the best and worst cases of random annotation by 14.85% and 26.04%, respectively. This demonstrates that valuable instance selection improves stability under extremely sparse annotations.

4.7 Visualization Analysis

To analyze samples selected by evaluation dimensions, we visualize the selected instances in Figure 3. Classification uncertainty selects visually ambiguous instances and higher classification difficulty, while localization uncertainty focuses on samples with larger box prediction deviations. Orientation uncertainty identifies instances with larger angle prediction errors. Inter-class diversity favors rare categories, while intra-class diversity selects visually distinctive instances. These results show that Active-SAOOD selects informative instances from corresponding perspectives. Furthermore, compared with models trained with random annotation, it reduces missed detections of distinctive and rare instances while improving orientation prediction for elongated objects.

5 Conclusion

In this work, we propose Active-SAOOD, an instance-level active learning framework addressing SAOOD performance degradation and instability under random sparse annotation. By considering instance-level evaluation dimensions for fundamental detection and representation generalization ability, Active-SAOOD selects high-value instances guided by model state observation strategy. Experiments demonstrate significantly improved performance and stability over random annotation under sparse annotation.

AI USE STATEMENT

In this work, we used generative AI tools to assist with translation, design and provide feedback on research methodology or experiments, and implement methods. We have not used generative AI tools to support qualitative and thematic data analysis, interpret results, or help develop theoretical models or conceptual frameworks, and the remaining required disclosure tasks, including generating synthetic datasets, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, and cleaning and reformatting datasets, are not applicable to this work. Additionally, we used generative AI tools for creating or editing software code, summarizing or analyzing existing literature, brainstorming, and editing the research paper to improve readability. We have reviewed all AI-assisted work. We checked AI-generated research ideas and literature analyses against relevant published literature through manual literature review, and AI-assisted code was manually inspected and tested for correctness. All AI-assisted manuscript content was reviewed and revised by the authors to ensure its accuracy and consistency. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Gong Cheng, Jiabao Wang, Ke Li, Xingxing Xie, Chunbo Lang, Yanqing Yao, and Junwei Han. Anchor-free oriented proposal generator for object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022.

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, 2021a.

Jiaming Han, Jian Ding, Nan Xue, and Gui-Song Xia. Redet: A rotation-equivariant detector for aerial object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2786–2795, 2021b.

Wenhao He, Xu-Yao Zhang, Fei Yin, and Cheng-Lin Liu. Deep direct regression for multi-oriented scene text detection. In *Proceedings of the IEEE international conference on computer vision*, pp. 745–753, 2017.

Wei Hua, Dingkang Liang, Jingyu Li, Xiaolong Liu, Zhikang Zou, Xiaoqing Ye, and Xiang Bai. Sood: Towards semi-supervised oriented object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15558–15567, 2023.

Dongyuan Li, Zhen Wang, Yankai Chen, Renhe Jiang, Weiping Ding, and Manabu Okumura. A survey on deep active learning: Recent advances and new frontiers. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4):5879–5899, 2024.

Yuxuan Li, Xiang Li, Yimain Dai, Qibin Hou, Li Liu, Yongxiang Liu, Ming-Ming Cheng, and Jian Yang. Lsknet: A foundation lightweight backbone for remote sensing: Y. li et al. *International Journal of Computer Vision*, 133(3):1410–1431, 2025.

Dingkang Liang, Wei Hua, Chunsheng Shi, Zhikang Zou, Xiaoqing Ye, and Xiang Bai. Sood++: Leveraging unlabeled data to boost oriented object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Wei Liao, Chunyan Xu, Chenxu Wang, and Zhen Cui. Llm-assisted semantic guidance for sparsely annotated remote sensing object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 22519–22528, 2025.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

Yu Lin, Jianghang Lin, Kai Ye, You Shen, Shengchuan Zhang, and Liujuan Cao. S^2teacher: Step-by-step teacher for sparsely annotated oriented object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 6997–7005, 2026.

Mingxin Liu, Peiyuan Zhang, Yuan Liu, Wei Zhang, Yue Zhou, Ning Liao, Ziyang Gong, Junwei Luo, Zhirui Wang, Yi Yu, et al. Partial weakly-supervised oriented object detection. *arXiv preprint arXiv:2507.02751*, 2025.

Zihan Lu, Chenxu Wang, Chunyan Xu, Xiangwei Zheng, and Zhen Cui. Progressive exploration-conformal learning for sparsely annotated object detection in aerial images. *Advances in Neural Information Processing Systems*, 37:40593–40614, 2024.

Jianqi Ma, Weiyuan Shao, Hao Ye, Li Wang, Hong Wang, Yingbin Zheng, and Xiangyang Xue. Arbitrary-oriented scene text detection via rotation proposals. *IEEE transactions on multimedia*, 20(11):3111–3122, 2018.

Peng Mi, Jianghang Lin, Yiyi Zhou, Yunhang Shen, Gen Luo, Xiaoshuai Sun, Liujuan Cao, Rongrong Fu, Qiang Xu, and Rongrong Ji. Active teacher for semi-supervised object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14482–14491, 2022.

Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9627–9636, 2019.

Keze Wang, Dongyu Zhang, Ya Li, Ruimao Zhang, and Liang Lin. Cost-effective active learning for deep image classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(12):2591–2600, 2016.

Xin Wang, Yudong Chen, and Wenwu Zhu. A survey on curriculum learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):4555–4576, 2021.

Yuting Wang, Velibor Ilic, Jiatong Li, Branislav Kisačanin, and Vladimir Pavlovic. Alwod: Active learning for weakly-supervised object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6459–6469, 2023.

Jiaxi Wu, Jiaxin Chen, and Di Huang. Entropy-based active learning for object detection with progressive diversity constraint. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9397–9406, 2022.

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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3974–3983, 2018.

Xingxing Xie, Gong Cheng, Jiabao Wang, Xiwen Yao, and Junwei Han. Oriented r-cnn for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3520–3529, 2021.

Hang Xu, Xinyuan Liu, Haonan Xu, Yike Ma, Zunjie Zhu, Chenggang Yan, and Feng Dai. Rethinking boundary discontinuity problem for oriented object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17406–17415, 2024.

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 *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3060–3069, 2021.

Chenhongyi Yang, Lichao Huang, and Elliot J Crowley. Plug and play active learning for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17784–17793, 2024.

Xue Yang, Junchi Yan, Ziming Feng, and Tao He. R3det: Refined single-stage detector with feature refinement for rotating object. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 3163–3171, 2021.

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. *arXiv preprint arXiv:2210.06742*, 2022a.

Xue Yang, Yue Zhou, Gefan Zhang, Jirui Yang, Wentao Wang, Junchi Yan, Xiaopeng Zhang, and Qi Tian. The kfiou loss for rotated object detection. *arXiv preprint arXiv:2201.12558*, 2022b.

Yi Yu and Feipeng Da. Phase-shifting coder: Predicting accurate orientation in oriented object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13354–13363, 2023.

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. *Advances in Neural Information Processing Systems*, 36:59137–59150, 2023.

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 *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 19283–19293, 2025a.

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*, 2025b.

Tianning Yuan, Fang Wan, Mengying Fu, Jianzhuang Liu, Songcen Xu, Xiangyang Ji, and Qixiang Ye. Multiple instance active learning for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5330–5339, 2021.

Ying Zeng, Yushi Chen, Xue Yang, Qingyun Li, and Junchi Yan. Ars-detr: Aspect ratio-sensitive detection transformer for aerial oriented object detection. *IEEE transactions on geoscience and remote sensing*, 62:1–15, 2024a.

Ying Zeng, Yushi Chen, Xue Yang, Qingyun Li, and Junchi Yan. Ars-detr: Aspect ratio-sensitive detection transformer for aerial oriented object detection. *IEEE transactions on geoscience and remote sensing*, 62:1–15, 2024b.

Ruixiang Zhang, Chang Xu, Haorao Zhu, Fang Xu, Wen Yang, Haijian Zhang, and Gui-Song Xia. Minimizing sample redundancy for label-efficient object detection in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

A APPENDIX

A.1 SUPPLEMENTARY MATERIAL OVERVIEW

- **Implementation Details** — Training settings, model configurations, loss functions, and annotation-budget matching. Sec. A.2
- **Hyperparameter Effects** — Effects of γ and β on instance selection and detection performance. Sec. A.3.1
- **Evaluation Dimension Ablation** — Contributions of classification, localization and orientation uncertainty, and inter-/intra-class diversity. Sec. A.3.2
- **MSO Generalization** — Generalization and rationality of model state observation using training and held-out data. Sec. A.3.3
- **MSO-guided Instance Difficulty** — How MSO adapts the difficulty of selected instances to the evolving model state. Sec. A.3.4
- **Cross-dataset Generalization** — Generalization of Active-SAOOD on DOTA-v1.5 and SODA-A. Sec. A.3.5
- **Annotation Cost Comparison** — Comparison with weakly/semi-supervised and point-/box-based annotation methods under equivalent annotation costs. Sec. A.3.6
- **Active Learning Iterations** — Effect of the number of active learning iterations on performance and training cost. Sec. A.3.7
- **Computational Cost** — Additional training cost introduced by LUP and MSO. Sec. A.3.8
- **Failure Cases** — Analysis of missed detections and misclassifications, including rare and heavily occluded objects. Sec. A.3.9
- **Visualization Analysis** — Visual comparison of randomly and actively selected instances, as well as qualitative detection results. Sec. A.4

A.2 MORE DETAILED EXPERIMENTAL SETTINGS.

Following most low-cost annotated oriented object detection studies Yang et al. (2022a); Hua et al. (2023), we adopt Rotated FCOS as our baseline detector, with a ResNet-50 backbone pretrained on ImageNet. Focal loss is used for classification with its parameters configured according to Lin et al. (2017). IoU loss is employed for regression, and a cross-entropy loss is applied to centerness. All loss weights are set to 1. Following previous sparsely annotation oriented object detection methods, we adopt a teacher-student framework to mine pseudo labels for sparse-annotation training. Following the common practice Hua et al. (2023), the EMA momentum of teacher model is set to 0.9996, and the teacher parameters are updated via EMA at each iteration. All remaining settings follow the configurations in mmrotate. To ensure a fair comparison, we strictly match the number of annotated objects between randomly sparse annotations and those actively selected by our method.

A.3 MORE DETAILED EXPERIMENTS.

A.3.1 MORE DETAILED HYPERPARAMETER EXPERIMENTS.

To analyze how the hyperparameters in our evaluation metrics influence both the selected instances and the final detection performance, we conduct detailed hyperparameter experiments. The hyperparameter γ in the inter-class diversity controls the slope of the sigmoid-like function. A larger γ yields a steeper slope, making the inter-class diversity more sensitive to the sampled number N_c. However, an excessively large γ causes the diversity score to drop too rapidly, weakening the contribution of this metric, while an overly small γ makes the diversity score insensitive to N_c, limiting its ability to highlight rare categories. As shown in Table 5, both overly large and overly small γ values degrade the effectiveness of the inter-class diversity, ultimately leading to performance drops in Active-SAOOD. When γ is set too small (e.g., 0.001), the inter-class diversity score is excessively large, which can mislead the overall evaluation of instance annotation value and result in training failure. When γ is set to 0.01, Active-SAOOD achieves the best performance, indicating that this value provides the most effective balance. Therefore, we set $\gamma = 0.01$ in experiments. We also conduct hyperparameter studies on β in the localization and orientation uncertainty metric. β controls the relative emphasis between orientation and localization uncertainties. In the extreme case where $\beta = 0$, $U_{\text{loc_}\theta}$ is determined solely by localization uncertainty. As β increases, the metric increas-

ingly prioritizes orientation uncertainty, particularly for objects with large aspect ratios. When β becomes excessively large, the orientation uncertainty for high–aspect-ratio objects dominates the metric, causing $U_{\text{loc_}\theta}$ to overlook localization uncertainty, which adversely affects performance. As shown in Table 5, setting $\beta = 0.5$ yields the highest performance. Therefore, we use $\beta = 0.5$ in experiments.

Table 5: Hyperparameter study. γ is the hyperparameter used in inter-class diversity, β is the hyperparameter used in localization and orientation uncertainty.

γ	AP(%)	β	AP(%)
0.001	-	0	29.31
0.005	31.96	0.25	33.97
0.01	**35.79**	0.5	**35.79**
0.05	33.15	0.75	31.80
0.1	22.62	1.0	30.26

A.3.2 Ablation study of evaluation dimensions for model ability enhancement.

To further demonstrate that the proposed evaluation dimensions improve the corresponding model capabilities during training, we conduct a fine-grained ablation study. As shown in Table 6, these dimensions can be categorized into fundamental detection ability and representation generalization ability. Specifically, classification uncertainty and localization and orientation uncertainty target fundamental detection ability, while inter-class and intra-class diversity enhance representation learning. When the classification uncertainty dimension is removed, the classification accuracy of the trained model drops to 76.94%, decreasing by 3.66% compared with the full model. This is because classification uncertainty prioritizes hard-to-classify instances, which directly improves the classification capability. Removing this dimension reduces the classification difficulty of selected samples, resulting in degraded performance. Similarly, removing localization and orientation uncertainty leads to a 10.65% decrease in average IoU, demonstrating its importance in enhancing the fundamental localization and orientation ability by selecting challenging instances. For representation generalization ability, removing inter-class diversity decreases rare-category accuracy by 3.34%. This is because inter-class diversity encourages the selection of rare categories and alleviates class imbalance in remote sensing images. Meanwhile, removing intra-class diversity reduces mean recall across categories, indicating degraded intra-class recall ability due to insufficient feature diversity learning under the large intra-class variations of remote sensing images. Overall, these results verify that the proposed instance-level evaluation dimensions can select samples that effectively improve the corresponding model abilities, thereby enhancing performance. This validates our motivation that instance annotation value lies in its potential contribution to improving the current detector.

A.3.3 Generalization and rationality of MSO model state observation.

Our MSO strategy guides Active-SAOOD in selecting the most suitable high-value samples by observing the model state on the initially annotated ground-truth (GT). To verify its generalization ability, we conduct corresponding experiments. To avoid data leakage caused by introducing validation or test data during training, the model state cannot be evaluated on validation or test sets to guide sample selection. Therefore, we study two settings for model state observation using MSO: (1) observated on the initially annotated GT set, and (2) observated on an independently held-out subset of comparable size that is randomly sampled and non-overlapping with the initially annotated GT set. In both cases, the observed model state is used to guide the selection of high-value instances for subsequent training.

As shown in Table 7, the final performance of Active-SAOOD is highly consistent regardless (less than 1% difference) of whether the model state is evaluated on the initial GT set or on the independent held-out set, and both settings yield significant performance improvements. This indicates that MSO is insensitive to the choice of data used as labels for model state evaluation, as MSO is designed to observe the relative evolution trend of the model's ability across different dimensions, rather than to estimate the absolute mAP in the conventional sense. Since this evolution reflects

Table 6: Ablation study of evaluation dimensions on model ability. All denotes the full model with all dimensions, while other rows with "w/o" remove the corresponding dimensions. Bold values indicate significant corresponding ability drops.

Setting	Fundamental Detection Ability		Representation Generalization Ability	
	Cls. Acc. (%)	Loc. IoU (%)	Rare Acc. (%)	Intra-class Recall (%)
All	80.60	69.77	69.86	60.30
w/o U_{cls}	**76.94**	69.06	68.56	59.26
w/o U_{loc_θ}	80.52	**59.12**	69.51	60.20
w/o D_{inter}	79.25	69.12	**66.52**	59.69
w/o D_{intra}	80.16	69.24	69.22	**55.68**



(a) (b)

Figure 4: Comparison with and without model state observation (MSO) strategy. (a) Training loss. (b) Training gradients. Traditional image-level active learning methods do not use MSO and directly select the most difficult instances, leading to unstable training. Active-SAOOD uses MSO to select appropriate instances according to the current model state, ensuring the training stability.

the intrinsic state of the model, its trend remains consistent when measured on either the training-participating initial GT or the non-training independent held-out set, leading to nearly identical final performance. Notably, under low annotation ratios (e.g., 1% annotation ratio), observing the model state on the initially annotated GT set even achieves slightly better performance than using an independent held-out set. This is mainly because, under extremely limited annotations, the model is insufficiently trained, and model state observation on a completely unseen independent held-out set may not accurately reflect the current model state. Consequently, the guided sample selection may become less suitable for the current model state. In addition, annotating an additional independent held-out set would incur extra labeling costs, we ultimately adopt the MSO on the initially annotated GT set for model state observation.

To further validate this observation, we evaluate the model ability on the evaluation set that is not involved in training. As shown in Table 8, the evolution trends of model ability observed on the initially labeled GT are consistent with those observed on non-training evaluation set. Specifically, during the early stage, the model abilities observed on both are relatively weak, causing MSO to favor relatively easy samples. As training progresses, the model abilities on both rapidly improve, leading MSO to gradually select challenging samples, and eventually converge in the late training stage. These results indicate that the sample selection guided by MSO based on the initially labeled GT is nearly identical to that guided by observations on non-training data, validating the generalization and rationality of MSO.

A.3.4 Experiments on instance difficulty selection guided by MSO.

To demonstrate the MSO strategy is able to select instances with appropriate difficulty according to the model capability at different training stages, we conduct the experiments. As shown in Table

Table 7: Generalization of the MSO strategy. MSO is employed to evaluate the model state on both the initially annotated GT and an independent held-out set of comparable size, respectively, to guide label selection.

Ability evalutation setting	Sparse annotation rate			
	1%	2%	5%	10%
Evaluated on initial GT	35.79	42.98	58.39	62.61
Evaluated on held-out set	34.96	42.67	58.32	62.60

Table 8: Experiments on the evolution trends of model ability across different training stages. Average ability is the mean of model ability dimensions in Eq. 11. The MSO strategy is employed to evaluate the current model ability on both the initially annotated GT set involved in training and the evaluation data excluded from training.

Annotation rate	Evaluation Set	Metric	Train stage			
			Early stage	Middle stage	Late stage	Convergence stage
1%	Initially labeled GT	Average Ability	0.49	0.63 (+0.14)	0.702 (+0.072)	0.705 (+0.003)
	Non-training evaluation set		0.40	0.49 (+0.09)	0.532 (+0.042)	0.534 (+0.002)
5%	Initially labeled GT	Average Ability	0.60	0.72 (+0.12)	0.750 (+0.030)	0.751 (+0.002)
	Non-training evaluation set		0.56	0.67 (+0.11)	0.698 (+0.028)	0.699 (+0.001)

8, at the early stage of training, the MSO strategy observes that the capabilities of model are relatively weak, and selects relatively simple instances. As training progresses, the capabilities of model rapidly improve, MSO correspondingly observes this enhancement and selects instances of increasing difficulty at the middle stage. At the late stage, MSO observes that the capabilities of model are already strong, and thus selects more difficult instances. Learning from these more challenging samples further boosts performance. Guided by MSO, selecting instances that are most suitable for the current model state, while gradually increasing the difficulty of the selected instances as the ability of model improves, effectively enhances the training stability and the overall performance.

As shown in Fig. 4, we compare the training loss of the traditional image-level active learning method, which directly selects the most difficult samples, with our state observation–based instance-level active learning approach that gradually increases the sample difficulty according to the model state. The traditional image-level active learning strategy tends to introduce a large number of hard samples at an early stage, which often leads to loss oscillation and unstable training in SAOOD. In contrast, the proposed model state observation guided active learning strategy progressively selects more difficult instances based on model state, achieving more stable training and better performance.

A.3.5 Generalization experiments on more datasets.

We further conduct generalization experiments of Active-SAOOD on the DOTA-v1.5 and SODA-A datasets, which contain large numbers of dense scenes and small objects. As shown in Table 9, Active-SAOOD consistently achieves significant performance improvements over S^2Teacher under different annotation ratios on both datasets, showing strong generalization capability of our method in dense scenarios.

Notably, the performance gains become even more significant under the extremely low annotation ratio of 1%. This is mainly because, under severely limited annotation budgets, random annotation is more likely to produce samples that are unsuitable for effective model learning, thereby wasting valuable annotation costs. In contrast, Active-SAOOD can actively select the most suitable high-value samples that best match the current model state, leading to substantial performance improvements under extremely low annotation ratios.

A.3.6 Comparison experiments with other reducing annotation cost methods.

We further conduct comparison experiments between Active-SAOOD and other reducing annotation costs methods. Since different settings adopt different annotation forms, we follow the relative annotation costs among RBox, HBox, and Point annotations provided in Wholly-WOOD Yu et al.

Table 9: Generalization experiments on different datasets.

Dataset	Method	Annotation Rate	
		1%	5%
DOTA-v1.5	Rotated FCOS	21.91	42.12
	S^2Teacher	27.61	47.72
	Active-SAOOD (S^2Teacher)	**33.72**	**50.76**
SODA-A	Rotated FCOS	37.54	59.56
	S^2Teacher	41.27	62.68
	Active-SAOOD (S^2Teacher)	**47.27**	**66.24**

Table 10: Comparison experiments with other reducing annotation cost methods.

Setting	Method	Label budget	mAP(%)
Weakly supervised (WS)	H2RBox-v2	150% Point	72.3
	Point2RBox-v2	100% Point	62.61
	Wholly-WOOD	100% Point	62.63
Partial WS	PWOOD	10% Point	42.35
Semi-supervised	SOOD++	20.5% Point	54.2
Active learning	Active-SAOOD	10.24% Point	58.39
		20.5% Point	62.6

(2025b) and convert all annotation budgets into an equivalent Point-based budget for a fair comparison. As shown in Table 10, under a 10% Equivalent point budget, Active-SAOOD reaches 58.39% AP, significantly outperforming partial weakly-supervised PWOOD Liu et al. (2025) (42.35%). Moreover, compared with the Point2RBox-v2 Yu et al. (2025a), Active-SAOOD achieves comparable performance with only one-fifth of the annotation budget. Although H2RBox-v2 Yu et al. (2023) (HBox supervised) achieves higher performance, its annotation budget is much higher than Active-SAOOD. Compared with weakly supervised methods, Active-SAOOD can achieve strong performance under extremely low annotation budgets, with Table 1 in the main paper highlight that Active-SAOOD achieves greater performance gains at lower annotation budgets, e.g., a 14.85% gain at 1% annotation rate and a 9.8% gain at 5%. Compared with the semi-supervised method SOOD++ Liang et al. (2025), Active-SAOOD achieves significantly higher performance under the same annotation budget (62.61% vs. 54.2%). Unlike weakly- or semi-supervised methods, which cannot select representative instances and suffer from feature redundancy and wasted annotation budget. The key advantage of Active-SAOOD lies in addressing this from a novel instance-level active learning perspective, enabling the active selection of high-value and representative instances, thereby achieving comparable or superior performance under extremely low annotation costs.

A.3.7 Experiment on the number of active learning iterations.

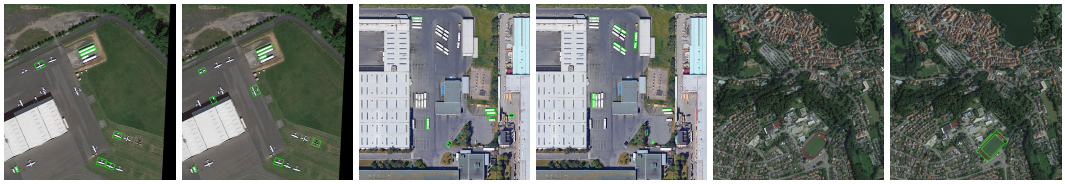
We further investigate the effect of the number of iterations in the active learning. As shown in Table 11, we set the number of iterations in Algorithm 1 to different values. When the number of active learning iterations increases from 2 to 3, the performance of Active-Rotated FCOS improves by 0.49%, while the performance of Active-SAOOD increases from 35.79% to 37.83%, yielding an improvement of 2.04%. This is mainly because a larger number of iterations corresponds to more active selection rounds. For example, increasing the iterations from 2 to 3 introduces an additional intermediate selection step. This allows Active-SAOOD to select more suitable samples based on the model state at intermediate training stages, thereby enhancing the detection performance. However, increasing the number of iterations also leads to higher training costs. Considering the trade-off between training cost and detection performance, we recommend setting the value of iterations in Algorithm 1 to 2.

Table 11: Effect of active learning iterations.

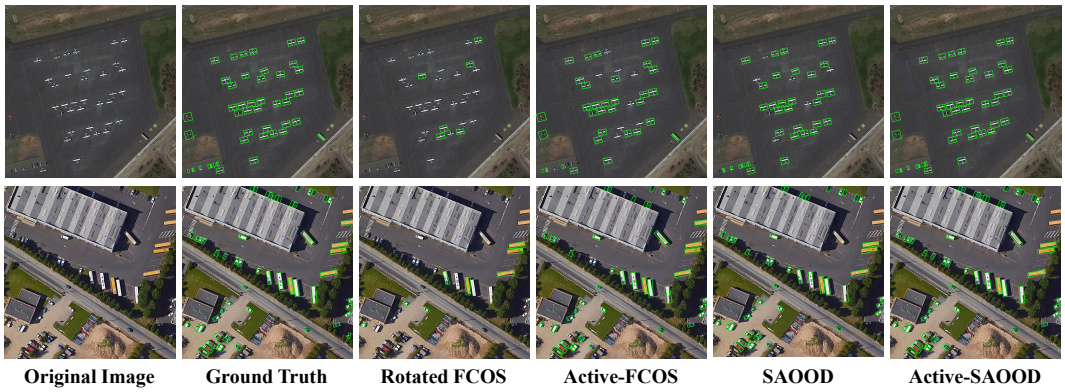
Method	Iteration number	mAP (%)
Active-Rotated FCOS	2	30.24
	3	30.73
Active-SAOOD	2	35.79
	3	37.83

Table 12: Computational cost of LUP and MSO.

Method	Train time (min)	Parameters (M)	Inference time (s)
+ LUP	2	0.0046	0
+ MSO	32	0	0
+ LUP + MSO	34	0.0046	0



(a)



(b)

Figure 5: Visualization analysis of instance selection and model performance. (a) Visualization comparing randomly annotated instances with actively selected instances. (b) Visualization comparing the detection results of models trained on randomly annotated instances versus those trained on actively selected instances.

A.3.8 Computational cost experiments of LUP and MSO.

Since our main goal is to reduce label costs while improving performance, we conduct a simple analysis of the additional computational costs introduced by LUP and MSO. Table 12 shows that both LUP and MSO are only applied during training and can be removed during inference, thus they do not affect inference speed. LUP adopts a simple 3×3 convolution, while MSO only involves evaluation metrics and softmax operations, resulting in negligible overhead in parameters and training time. These results show that Active-SAOOD does not introduce significant overhead.

A.3.9 Failure cases analysis.

To gain a deeper understanding of Active-SAOOD, we further conduct a visualization analysis of several failure cases during detection. As shown in Fig. 6a, for small vehicles heavily occluded by shadows, the extremely blurred appearance still leads to missed detections. However, we observe that even under such extreme conditions, one small vehicle is still correctly detected. This is mainly because the classification uncertainty encourages the selection of classification-challenging instances for training, which in turn helps the model better recognize such difficult objects. Furthermore, Active-SAOOD occasionally misclassifies truck containers as large vehicles, likely because the increased intra-class diversity enriches the feature space of the training set, and the visual similarity between truck containers and large vehicles is very high, making them harder to distinguish. As shown in Fig. 6b, for extremely rare categories such as helicopter, although the inter-class di-



(a) (b)

Figure 6: Visualization of failure cases. Red boxes are false positives, green are true positives.

versity prioritizes selecting these instances for annotation, the number of rare-category samples in the training set remains limited under very low annotation ratios. As a result, some rare objects may still be missed detection. Nevertheless, compared with random sparse annotation, the detection accuracy for rare categories is already significantly improved. In addition, during visualization, we observed that some small vehicles partially occluded in the images were overlooked during the manual annotation. Owing to the improved generalization ability brought by actively selected instances, Active-SAOOD is still able to detect these manually missed small vehicles, which further demonstrates the strong generalization ability of our Active-SAOOD.

A.4 More detailed visualization analysis.

To deeper insights into the high-value instances selected by Active-SAOOD, we conduct a comparative visualization analysis between randomly selected instances and those actively selected by our method. As shown in Fig. 5a, random selection exhibits strong randomness, and under low annotation ratios it often selects instances with highly similar characteristics, thereby reducing the overall feature diversity of the annotated data. For example, in the left figure, most selected planes share very similar appearances, while more visually distinctive airplanes remain unannotated. In contrast, Active-SAOOD leverages our proposed intra-class diversity metric to evaluate and select instances with more distinctive characteristics, ensuring richer feature diversity even under low annotation ratios. Moreover, compared with random selection, Active-SAOOD tends to actively choose objects with large aspect ratios (such as the large-vehicle in the middle figure). This is because orientation prediction errors have a stronger impact on the localization of such objects, and our localization and orientation uncertainty metric encourages selecting these challenging instances for training. For rare categories (such as the runway and sports field in the right figure), random selection often fails to include them under low annotation ratios. Active-SAOOD, however, prioritizes these rare categories through our inter-class diversity metric, preventing model overfitting and improving detection accuracy for infrequent classes.

To understand how training with actively selected instances benefits the detector, we compare the detection results of models trained on different instance selection strategies. As shown in Fig. 5b, Rotated FCOS represents the baseline trained with randomly selected sparse annotations, while Active-FCOS denotes the same baseline trained with instances actively selected by our method (with the annotation budget kept identical). Both models share the same network architecture and training settings. SAOOD is trained on randomly selected sparse annotations but incorporates pseudo-label mining, whereas Active-SAOOD uses actively selected instances together with pseudo-label.

Due to sparse supervision, Rotated FCOS lacks sufficient positive samples, especially under low annotation ratios. The limited diversity of the annotated instances causes the model to overfit easily, resulting in a small number of detected objects. In contrast, Active-FCOS benefits from our actively selected high-value samples, which provide richer feature diversity even under low annotation ratios. This alleviates overfitting to some extent, enabling the model to detect more objects. However, without pseudo-label training, unlabeled objects still confusion foreground and background features, leading to missed detections, particularly for small vehicles. SAOOD alleviates overfitting by mining pseudo-labels for unlabeled objects, thereby increasing the number of positive samples and reducing missed detections. Nevertheless, the localization and orientation accuracy of pseudo-labels is not always reliable. This can negatively impact the ability of the model to accurately

localize objects and predict their orientations, causing angle prediction errors for large aspect-ratio objects (e.g., large-vehicle). Active-SAOOD addresses this issue by selecting localization and orientation challenging objects through our proposed evaluation dimension, allowing the model to better learn more accurate localization and orientation for difficult cases. As a result, large aspect-ratio objects with challenging rotation angles exhibit more precise orientation predictions. For example, the yellow large-vehicle in Fig. 5b shows notably improved angle prediction accuracy. In addition, objects with distinctive features, as well as partially occluded objects, are better detected because our method selects high value instances using classification uncertainty and intra-class diversity, further enhancing the overall detection performance.