

A Simple Baseline with Placement Prior for Point-Supervised Oriented Object Detection

Runxiang Liu[✉], Kaikai Xie[✉], Qianxi Cao[✉],
Yuming Fang[✉], Jiebin Yan[✉], and Junjie Chen^{*✉}

Jiangxi University of Finance and Economics, Nanchang, China
{2202305238, 2202405677, 2202426517}@stu.jxufe.edu.cn
{fangyuming, yanjiebin, chenjunjie}@jxufe.edu.cn

Abstract. Oriented object detection for aerial images has attracted growing attention due to the increasing demand in extensive applications. However, the oriented box annotation is relatively expensive, and thus point-supervised oriented object detection has emerged as a practical paradigm. Owing to the lack of size and angle annotations, existing methods leverage various complementary information, but usually suffer from extra training stages or multiplied computational overhead. In this paper, we propose three simple yet effective loss terms, and set a simple baseline with the placement prior. In particular, the placement prior indicates that objects in aerial images are placed in a context, rather than exhibiting arbitrary distribution. Therefore, we propose the angle prior loss, considering that the angles of objects are usually aligned in a neighborhood. Furthermore, we propose the size prior loss, considering that the sizes of objects are usually proportional for each class in an image. Besides, we propose to derive oriented boxes from the Voronoi Watershed regions without training, which can jointly regularize the estimated angles and sizes. By employing only loss terms, our baseline is not only simple and easy to extend but also readily transferable to other methods. We conduct extensive experiments on representative UAV datasets and remote sensing datasets, which indicates the effectiveness of our method.

Keywords: Oriented Object Detection · Point Supervision

1 Introduction

Oriented object detection for aerial images plays a vital role in various practical applications, such as aerial scene analysis [9, 28, 43] and autonomous driving [8]. However, annotating oriented bounding boxes is relatively costly and labor-intensive, as each object requires labeling multiple parameters. To alleviate annotation costs, point-supervised oriented object detection has garnered increasing attention for its low annotation burden, *i.e.*, only a single point and its corresponding category label per object. This paradigm provides a practical and cost-effective solution for training accurate oriented object detectors.

* Corresponding author.



Fig. 1: The illustration of the placement prior in representative UAV and remote sensing datasets. Generally, the target objects are placed on the ground plane in a context, rather than exhibiting arbitrary distribution.

Given only point annotations, the major issue of point-supervised oriented object detection is to exploit various extra information to acquire box sizes and orientations. To name a few, PointOBB-v2 [36] pre-trains to generate class probability maps to derive pseudo rotated bounding boxes (RBoxes), and then trains the detector with these RBoxes. PointOBB-v3 [62] constructs multiple transformed (augmented) views of the original image, and learns RBoxes by scale augmentation module, angle acquisition module, and consistency loss. Point2RBox-v2 [58] regularizes the estimated height and width via Voronoi watershed loss, and learns size/rotation variation via consistency loss on transformed views. Recently, Point2RBox-v3 [64] combines the watershed algorithm with the Segment Anything Model (SAM) for dynamic pseudo-label refinement. While the above works have achieved significant advances, they are relatively complex and heavy, *e.g.*, pseudo labeling requires extra training stages, and duplicated transformations require multiplied computational overhead. In contrast, a simple and compact baseline could lower research barriers and facilitate further exploration.

In this paper, we propose three simple yet effective loss terms, and set a simple baseline based on the placement prior of aerial images, termed PlacePriorBase. In particular, the placement prior implies that objects in aerial images are placed in a context, rather than exhibiting arbitrary distribution. As shown in Fig. 1, the target objects placed in a neighborhood are generally aligned (*i.e.*, perpendicular or parallel to each other), and the target objects of the same class are generally in proportional sizes due to the placement on the ground plane. Motivated by such observation, we design angle prior loss to regularize the estimated rotation angles of positive boxes in neighborhoods, and design

size prior loss to regularize the estimated width and height of positive boxes of the same class. Notably, all geometric attributes constrained within these priors are real-time network predictions rather than ground-truth bounding boxes, thereby strictly maintaining the single-point supervised paradigm. Even though the placement prior is not a strictly valid principle, it can effectively provide beneficial prior knowledge to regularize the rotation angle and object size under the point-supervised setting.

Besides, we propose a simple yet effective strategy to derive RBox targets based on Voronoi Watershed regions [1, 41], inspired by [58]. Specifically, we derive a prior angle from the enclosing RBox on the default Voronoi Watershed regions, and then expand to a better RBox by assigning directional priorities during watering. In contrast to [58] that relies on trained angle estimations, our strategy is not only training-free but also delivers extra angle supervision.

Overall, our proposed loss terms are complementary to most existing works and can be easily integrated. We conduct comprehensive experiments on four representative datasets of aerial images, including two Unmanned Aerial Vehicle (UAV) datasets (*i.e.*, DroneVehicle [38] and CODrone [56]) and two remote sensing datasets (*i.e.*, DOTA-v1.0 [43] and DOTA-v1.5, an extended version based on [43]). The extensive comparison and in-depth analysis demonstrate the effectiveness of our proposed loss terms and PlacePriorBase. Our main contributions can be summarized as follows:

- (1) We set a simple baseline, namely PlacePriorBase, with the placement prior, which indicates that objects are placed in a natural context from a bird’s-eye view.
- (2) We propose the size prior loss, angle prior loss, and Voronoi Watershed RBox loss, which are standalone and easily transferred to other methods.
- (3) We conduct extensive experiments to show the performance of PlacePriorBase on various benchmarks and the complementarity of our loss terms on other methods.

2 Related Works

Fully-Supervised Oriented Object Detection Fully-supervised oriented object detection takes high-precision rotated bounding box (RBox) annotations as supervision signals, aiming to achieve accurate localization and classification of objects with arbitrary orientations, and it is the most mature technical direction in this field. Current methods fall into several categories: anchor-based frameworks like Rotated RetinaNet [25], anchor-free designs such as Rotated FCOS [40], and two-stage architectures including RoI Transformer [7], Oriented R-CNN [44], and ReDet [12]. To improve feature adaptation for rotated instances, alignment-based detectors like R3Det [34] and S²A-Net [11] have been proposed.

Angle regression in OOD suffers from boundary discontinuity, which various strategies aim to resolve. These include angle encoding schemes [45–47, 57] that map angles to continuous representations, refined loss functions [35, 48, 50] that mitigate optimization instability, and Gaussian-based formulations [33, 49, 51, 53,

54] that treat rotated boxes as probabilistic distributions. Alternative paradigms also exist: RepPoint-based methods [14, 21, 55] represent objects via adaptive point sets, while LMMRotate [20] leverages multimodal language models for autoregressive localization.

Despite these advances, most methods depend on fully annotated oriented boxes, which are costly to obtain. Reducing this annotation burden while effectively utilizing weaker annotations remains an open challenge. The acquisition of such annotations requires extremely high labor and time costs, making it difficult to adapt to the practical application requirements of rapid deployment.

Point-Supervised Oriented Detection Point-supervised oriented detection has recently achieved various representative advances. Several methods [2, 3, 15, 16, 23, 27, 63] utilize the zero-shot segmentation capability of SAM [19] to convert point prompts into oriented boxes. Point2RBox [59] takes a different technical route, constructing an end-to-end detector through knowledge integration, though it requires one-shot examples per target category. Its successors, Point2RBox-v2 [58] and Point2RBox-v3 [64], further incorporate Voronoi watershed segmentation and SAM-generated priors respectively to improve accuracy. The PointOBB series [31, 36, 62] adopts a prior-free strategy: the original PointOBB [31, 62] generates rotated boxes via scale consistency and multiple instance learning. PointOBB-v2 [36] introduces class probability maps to produce higher-quality pseudo labels with a simpler pipeline. PointOBB-v3 [62] pushes performance boundaries further through refined training strategies and better proposal initialization. Beyond point-to-box conversion, GenSAM [15] learns task-level generic prompts for robust promptable segmentation, while ProMaC [16] further improves segmentation reliability by verifying and correcting hallucinated prompt-induced regions.

While these methods have significantly advanced point-supervised oriented detection, they often introduce considerable complexity. Many methods depend on multi-view transformations or duplicated computations that multiply training costs. The field still lacks a simple yet effective baseline that achieves strong performance through lightweight, compact designs without complex pipelines or external dependencies.

Other Weakly-Supervised Settings Beyond point-supervised oriented detection, other weakly supervised paradigms have also seen substantial progress. The “Point-to-HBox” task is most directly related to point-supervised oriented detection. PSOD [10] combines edge detection with adaptive masked flood fill to localize salient objects from point annotations. P2BNet [6] samples and classifies box proposals around annotated points for point-supervised horizontal detection. The “HBox-to-RBox” task generates rotated boxes from horizontal annotations. H2RBox [52] applies geometric constraints to limit object angles and uses a self-supervised branch to refine predictions. H2RBox-v2 [60] further leverages reflection symmetry for improved accuracy. EIE-Det [42] introduces explicit and implicit equivariance branches to learn rotation, position, and scale consistency. KCR [66] explores transfer learning by jointly using rotated and

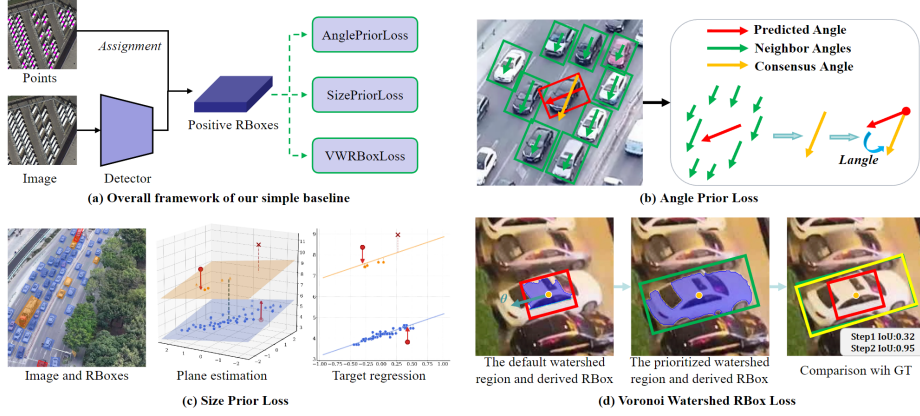


Fig. 2: The overview of our simple baseline and proposed loss terms. (a) PlacePriorBase only includes loss terms on the basic detector. (b) Our angle prior loss encourages the angle alignment in neighborhoods. (c) Our size prior loss encourages linear size estimations for objects in the same class. (d) Our Voronoi Watershed RBox loss finds a prior angle and derives a better RBox for supervision.

horizontal box datasets. Some methods [17, 39] incorporate extra annotations for better performance, though their generalizability is constrained by annotation cost. The “Point-to-Mask” task employs large-scale pre-trained models to enable instance mask generation from point annotations. Point2Mask [22] designs a single-point annotation scheme for panoptic segmentation. Although oriented boxes can be derived from masks via minimum enclosing rectangles, this cascaded pipeline (Point-to-Mask-to-RBox) is computationally inefficient and underperforms end-to-end methods, further motivating the need for dedicated simple baselines [52, 60]. Beyond single-point annotations, recent advanced methods [4, 5, 26, 30, 61, 65] explore relaxed or hybrid weakly-supervised paradigms. For instance, Wholly-WOOD [61] wholly leverages diversified-quality labels. Furthermore, PWOOD [26] and SPWOOD [65] formulate partial or sparse partial weakly-supervised frameworks to optimize oriented object bounding boxes from incomplete supervisory signals.

3 Method

3.1 Task Definition

In the training stage, each training sample consists of the RGB image $\mathbf{x} \in \mathbb{R}^{H \times W \times 3}$ and N object annotations $\{\mathbf{p}_i^*, \mathbf{y}_i^*\}_{i=1}^N$. For the i -th object, $\mathbf{p}_i^* \in \mathbb{R}^2$ indicates the 2D point coordinates, and $\mathbf{y}_i^* \in \mathbb{R}^C$ represents the one-hot label for C object classes. In the test stage, the detector should estimate the oriented box (*i.e.*, (x, y, w, h, θ)) and class label (*i.e.*, $\mathbf{y}$) for each object in the input image.

3.2 Method Overview

Our loss terms could be well applied to various detection models. To build PlacePriorBase, we follow the representative Point2RBox-v2 [58] and employ a simple architecture with ResNet50 [13] backbone, FPN [24] head, and PSC [57] angle coder. As shown in Fig. 2 (a), for the network estimations, our loss terms provide standalone, compact, and targeted supervisions. Specifically, given an input image $\mathbf{x}$, the model extracts feature maps via backbone and FPN, and finally produces pixel-wise RBoxes on the P3 layer. Each RBox consists of oriented box (*i.e.*, (x, y, w, h, θ)) and class label (*i.e.*, $\mathbf{y}$). For point assignment, the estimated RBoxes are assigned to one point annotation, if the estimation pixel is in the neighborhood of that point. For brevity, we refer to the estimated RBoxes assigned with point annotations as positive RBoxes, and the remaining ones as negative RBoxes. Overall, we will supervise the positive RBoxes with angle prior loss, size prior loss, Voronoi Watershed RBox loss, and the basic center loss, and supervise all estimated class labels with classification loss.

3.3 Angle Prior Loss

As shown in Fig. 2 (b), our angle prior loss encourages the angle alignment of positive RBoxes in the neighborhoods. Assume that there are M positive RBoxes in the current training sample, and the i -th positive RBox is denoted as $(x^{(i)}, y^{(i)}, w^{(i)}, h^{(i)}, \theta^{(i)})$. Specifically, for the i -th positive RBox, we compute the aligned angle $\mathcal{A}(i)$ for its neighborhood by:

$$\mathcal{A}(i) = \frac{\sum_{j \neq i} \mathcal{W}(i, j) \cdot \text{vec}(\theta^{(j)})}{\sum_{j \neq i} \mathcal{W}(i, j)}, \quad (1)$$

where $[\cdot]$ indicates the tensor concatenation, $\text{vec}(\theta) = [\cos(4\theta); \sin(4\theta)]$ vectorizes the angle θ , and $\mathcal{W}(i, j)$ is the neighborhood weight controlling the influence from the j -th RBox to the i -th RBox. In particular, the angle vectorization ensures that the angle discrepancy is measured in range $[0, \frac{\pi}{4})$, considering that objects could be aligned in parallel or perpendicular. The neighborhood weight is defined as:

$$\mathcal{W}(i, j) = \exp\left(\frac{-\|x^{(i)} - x^{(j)}\| - \|y^{(i)} - y^{(j)}\|}{2 \cdot w^{(i)} \cdot h^{(i)}}\right), \quad (2)$$

which is scaled by the center distance from the j -th RBox to the i -th RBox. Finally, we optimize the negative align energy of all neighborhoods, as:

$$\mathcal{L}_{\text{angle}} = \frac{1}{M} \sum_i (1 - \text{vec}(\theta^{(i)}) \cdot \frac{\mathcal{A}(i)}{\|\mathcal{A}(i)\| + \epsilon}), \quad (3)$$

where the scalar 1 ensures a positive loss value, and $\epsilon = 0.01$ avoids division by zero in unit vector normalization.

3.4 Size Prior Loss

As shown in Fig. 2 (c), our size prior loss encourages the proportional size of positive RBoxes for the same class in each image. Due to lacking size annotations, we assume that objects of the same class have similar sizes, and thus the sizes are approximately proportional when placed on the ground plane and captured from some perspective. In this way, an approximated size could be solved for each instance to complement the missing of size supervision.

Specifically, we assume that there are M positive RBoxes, and the i -th positive RBox of the c -th object class is denoted as $(x^{(i,c)}, y^{(i,c)}, w^{(i,c)}, h^{(i,c)}, \theta^{(i,c)})$. Based on the principle of perspective, the position and size of a single object generally satisfy:

$$k_x \cdot x^{(i,c)} + k_y \cdot y^{(i,c)} + b_c = w^{(i,c)}, \quad (4)$$

where (k_x, k_y) is the gradient of the ground plane, and b_c is the size intercept of the c -th class. b_c could be understood as the basic size at original point, and the approximated size exhibits a linear variation with respect to positional changes. For conciseness, we only describe the approximation and loss for the width in Eqn. 4 and Eqn. 7, which are equivalently applied for height.

Considering that there are C classes and the ground plane is shared by all objects, we have $C + 2$ parameters $\mathbf{W} = [k_x; k_y; b_1; b_2; \dots; b_C]$ to solve for each image. Therefore, we employ ridge regression to solve $\mathbf{W}$ for each image:

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} \|\mathbf{X}\mathbf{W} - \mathbf{Y}\|^2 + \lambda \|\mathbf{W}\|^2. \quad (5)$$

In particular, each row of $\mathbf{X}$ corresponds to the input part of Eqn. 4 for one object, and the row of the i -th positive RBox of the c -th class is:

$$\mathbf{X}_{i,c} = [x^{(i,c)}; y^{(i,c)}; \mathbb{I}_{c=1}; \mathbb{I}_{c=2}; \dots; \mathbb{I}_{c=C}], \quad (6)$$

where $\mathbb{I}_{c=1}$ is the indicator function taking the value of 1 if $c = 1$, and 0 otherwise. Each row of $\mathbf{Y}$ corresponds to the output part of Eqn. 4, and the row of the i -th positive RBox of the c -th class is $\mathbf{Y}_{i,c} = w^{(i,c)}$. Via `torch.linalg.solve`, we could effectively solve the optimal $\hat{\mathbf{W}}$ and employ it to obtain a target size for regularizing the positive RBoxes:

$$\mathcal{L}_{\text{size}} = \|\mathbf{X}\hat{\mathbf{W}} - \mathbf{Y}\|^2. \quad (7)$$

3.5 Voronoi Watershed RBox Loss

A Voronoi diagram [1] is a partitioning of a space based on a set of points, and thus point annotations could be naturally employed to partition an exclusive region indicating the raw shapes of objects. Specifically, the exclusive region is obtained by watershed with point annotations as foreground markers and Voronoi ridges as background boundaries (termed as watering), *i.e.*, $\{\mathbf{S}_i\}_{i=1}^N = \text{Watershed}(\mathbf{x}, \{\mathbf{p}_i^*\}_{i=1}^N)$, where $\mathbf{S}_i$ is the output basin region (pixel coordinates) of the i -th point annotation. In [58], the target width and height of i -th point

annotation based on $\mathbf{S}_i$ are obtained depending on the model’s on-line angle estimation, which requires comprehensive model training and may be misled by immature angle estimations.

As shown in Fig. 2 (d), since the basin region depicts a raw shape, we propose to find a prior angle $\hat{\theta}_i$ according to the enclosing RBox of $\mathbf{S}_i$. Furthermore, based on the prior knowledge of angle $\hat{\theta}_i$, we obtain better basin regions by assigning directional priorities during watering:

$$\{\mathbf{S}'_i\}_{i=1}^N = \text{PrioritizedWatershed}(\mathbf{x}, \{\mathbf{p}_i^*\}_{i=1}^N, \{\hat{\theta}_i\}_{i=1}^N). \quad (8)$$

The details of PrioritizedWatershed are left to Supplementary. Then, we derive the target width and height from the enclosing RBox of $\mathbf{S}'_i$, with prior angle $\hat{\theta}_i$ and the center point $\mathbf{p}_i^*$:

$$[\hat{h}_i; \hat{w}_i] = \text{EnclosingBox}(\mathbf{S}'_i, \mathbf{p}_i^*, \hat{\theta}_i). \quad (9)$$

Finally, we supervise the positive RBox assigned to point $\mathbf{p}_i^*$ by:

$$\mathcal{L}_{\text{RBox}} = \frac{1}{M} \sum_{i=1}^M \mathcal{L}_{\text{GWD}} \left(\mathbf{R}^{(i)} \begin{bmatrix} \frac{w^{(i)}}{2} & 0 \\ 0 & \frac{h^{(i)}}{2} \end{bmatrix}^2, \hat{\mathbf{R}}_i \begin{bmatrix} \frac{\hat{w}_i}{2} & 0 \\ 0 & \frac{\hat{h}_i}{2} \end{bmatrix}^2 \hat{\mathbf{R}}_i^\top \right) \quad (10)$$

where $\mathcal{L}_{\text{GWD}}(\cdot)$ is the Gaussian Wasserstein Distance Loss [53], and $\mathbf{R}^{(i)}$ (similarly $\hat{\mathbf{R}}_i$) denotes the 2D rotation matrix parameterized by the predicted angle $\theta^{(i)}$ (or the prior angle $\hat{\theta}_i$) as:

$$\mathbf{R}^{(i)} = \begin{bmatrix} \cos \theta^{(i)} & -\sin \theta^{(i)} \\ \sin \theta^{(i)} & \cos \theta^{(i)} \end{bmatrix}. \quad (11)$$

Compared with previous work [58] that only supervises width and height, our loss term can supervise the entire rotated box and requires no training.

3.6 Training and Inference

The overall training of PlacePriorBase is quite compact and clear, which contains a single stage without duplicated computations. The full training objective is:

$$\mathcal{L}_{\text{full}} = \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{center}} + \alpha \mathcal{L}_{\text{angle}} + \beta \mathcal{L}_{\text{size}} + \gamma \mathcal{L}_{\text{RBox}}, \quad (12)$$

where $\alpha = 0.2$, $\beta = 0.1$, and $\gamma = 0.5$ are the balancing hyper-parameters for our proposed loss terms. Since our loss terms are not strong supervision from full annotations, we add slacks by omitting the top-10% loss values, similar to the slack implementation in [58]. $\mathcal{L}_{\text{cls}}$ and $\mathcal{L}_{\text{center}}$ are the default classification loss and center point loss, which supervise the estimated class labels and push the estimated RBox centers to the given point annotations. During inference, the pipeline is consistent with the original architecture in [58]. Crucially, our proposed prior loss terms only serve as auxiliary constraints during the training phase and are completely discarded in inference.

4 Experiments

4.1 Datasets, Metrics and Implementation Details

Datasets We conduct experiments on representative UAV datasets and remote sensing datasets for a comprehensive validation on aerial images. Specifically, DroneVehicle [38]. A large-scale RGB-thermal dual-modal dataset collected via UAVs, containing 28,439 pairs of visible-infrared images with 5 vehicle categories. We only use the visible spectrum subset in this work for a fair comparison with all counterpart methods. We train our model on the training set and test it on the test set.

CODrone [56]. A UAV-based remote sensing dataset for oriented object detection, comprising 10,004 high-resolution aerial images annotated with bounding polygons for 12 urban target classes. It covers diverse complex urban scenarios with significant variations in scale, orientation, and distribution. All experiments use its training and validation sets for training and test set for evaluation.

DOTA [43]. DOTA-v1.0 consists of 2,806 high-resolution aerial images with annotations for 15 common object categories. DOTA-v1.5 keeps the same set of images but adds one more object category, making the annotation settings more comprehensive for oriented object detection. In our experiments, we follow the official data splits for both datasets and conduct all model training on the training and validation sets, with performance evaluated on the test set.

Metrics Following previous works [58, 59, 62], we adopt mAP@0.5 as the evaluation metric for detection performance, which is defined as the mean Average Precision across all object categories where the Intersection over Union (IoU) between predicted RBoxes and their GTs is set to 0.5.

Implementation Details Our implementation builds on Point2RBox-v2 [58], which is developed in PyTorch and MMRotate. We adopt the classic FCOS [40] with a ResNet-50 [13] backbone as the detector, and employ the PSC angle coder for rotation encoding. All models are trained using AdamW [29] with an initial learning rate of 5×10^{-5} , a warm-up of 500 iterations, and a decay factor of 10 at each scheduled step. We train for 12 epochs on all datasets, applying random flip as the only data augmentation. All experiments are conducted on NVIDIA RTX 4090 GPU and follow the same hyper-parameters. For data preprocessing, images from DOTA and CODrone are cropped into 1024×1024 patches with a 200-pixel overlap, while DroneVehicle images are used at their original 800×800 resolution. All experiments are evaluated without multi-scale techniques.

4.2 Comparison with Previous Methods

Baseline Setting We set two groups of comparable baselines: (1) RBox-supervised methods, which learn from the full annotations of RBoxes and show the upper-bound against point-supervised setting, including YOLOv5s [18], Faster R-CNN [37], FCOS [40], RTMDet [32], S²A-Net [11]. (2) Point-supervised methods,

Table 1: The quantitative comparison of various methods on four aerial datasets.

Method	*	DroneVehicle	CODrone	DOTA-v1.0	DOTA-v1.5
▼ RBox-supervised OOD					
YOLOv5s (2020) [18]	✓	72.29	45.99	76.89	73.19
Faster R-CNN (2015) [37]	✓	60.06	40.40	73.40	62.00
FCOS (2019) [40]	✓	65.73	44.70	72.44	64.53
RTMDet (2022) [32]	✓	73.57	46.31	80.54	78.12
S ² A-Net (2022) [11]	✓	65.96	37.61	75.81	66.53
▼ Point-supervised OOD					
PointOBB (2024) [31]	×	27.58	12.22	30.08	10.66
PointOBB-v2 (2025) [36]	×	40.24	18.11	41.68	30.59
PointOBB-v3 (2025) [62]	✓	42.43	18.27	41.29	31.25
PointOBB-v3 (2025) [62]	×	51.11	18.85	49.24	33.79
Point2RBox (2024) [59]	✓	40.61	17.53	40.27	30.51
Point2RBox-v2 (2025) [58]	✓	53.70	22.30	51.00	39.45
Point2RBox-v2 (2025) [58]	×	63.49	27.53	62.61	54.06
Point2RBox-v3 (2025) [64]	✓	61.86	28.42	59.61	48.34
Point2RBox-v3 (2025) [64]	×	66.77	29.37	66.09	56.86
▼ Point-supervised OOD (Ours)					
PlacePriorBase	✓	40.75	21.70	42.05	32.38
▼ Point-supervised OOD (SOTA+Ours)					
PlacePrior+POBB3	✓	44.37	19.38	42.34	32.88
PlacePrior+POBB3	×	54.95	21.44	50.95	35.97
PlacePrior+P2R2	✓	56.37	24.11	52.63	43.06
PlacePrior+P2R2	×	65.38	29.12	63.92	55.47
PlacePrior+P2R3	✓	64.15	30.43	61.09	51.61
PlacePrior+P2R3	×	68.49	31.06	67.29	58.15

*Comparison tracks: ✓ = End-to-end training and testing; × = Generating pseudo labels to train the FCOS detector (two-stage training).

which learn from only point annotations of RBoxes, including PointOBB [31], PointOBB-v2 [36], PointOBB-v3 [62], Point2RBox [59], Point2RBox-v2 [58]. Our method falls into the point-supervised group, and we include several variants. Specifically, the basic version is PlacePriorBase introduced in Sec. 3. Further, we integrate our loss terms into representative SOTA baselines, including PointOBB-v3 [62], Point2RBox-v2 [58], and Point2RBox-v3 [64], yielding PlacePrior+POBB3, PlacePrior+P2R2, and PlacePrior+P2R3 respectively.

Quantitative Results The results of various methods on four datasets are summarized in Tab. 1, where we have the following observations. Firstly, PlacePriorBase achieves satisfactory and cost-effective performances against compared baselines. *E.g.*, PlacePriorBase achieves on average 82.45% of the end-to-end version of Point2RBox-v2, and outperforms most baselines (+3.59% and +1.79% against PointOBB-v2 on CODrone and DOTA-v1.5). Secondly, when our loss

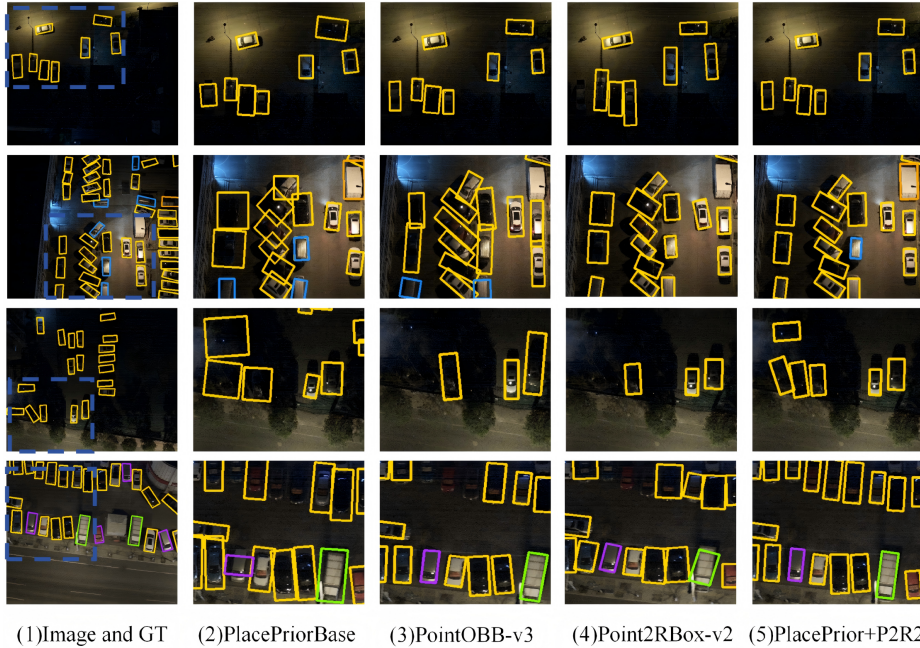


Fig. 3: The qualitative comparison against competitors on the DroneVehicle dataset. The dashed boxes in col. (1) indicate enlarged regions for clearer comparisons.

terms are applied to the SOTA baselines, the performances are further improved, demonstrating that our losses generally provide complementary and beneficial information. *E.g.*, PlacePrior+P2R2 achieves the optimal performance against all baselines, (+1.81% and +3.61% against Point2RBox-v2 on CODrone and DOTA-v1.5). Moreover, the recent Point2RBox-v3 [64] demonstrates outstanding performance, and integrating our loss terms (PlacePrior+P2R3) yields further significant improvements. Therefore, PlacePriorBase is cost-effective for both UAV and remote sensing datasets, and our loss terms could be effectively transferred to other methods.

Qualitative Results We select end-to-end and representative baselines and our variants for qualitative comparison, including PlacePriorBase, PointOBB-v3 [62], Point2RBox-v2 [58], and PlacePrior+P2R2. We show the visualizations on DroneVehicle dataset in Fig. 3 and show the visualizations on DOTA-v1.0 dataset in Fig. 4. Due to the lack of GT annotations on the test set of DOTA-v1.0, we employ models trained on training set and visualize on val set. Overall, we have the following observations. Firstly, PlacePriorBase generally yields angle-aligned, size-consistent, and relatively satisfactory RBox estimations. When integrated into Point2RBox-v2, PlacePrior+P2R2 presents clear improvements, demonstrating that our loss terms could be effectively transferred to the SOTA baseline and provide complementary information. *E.g.*, as shown in the 2nd row

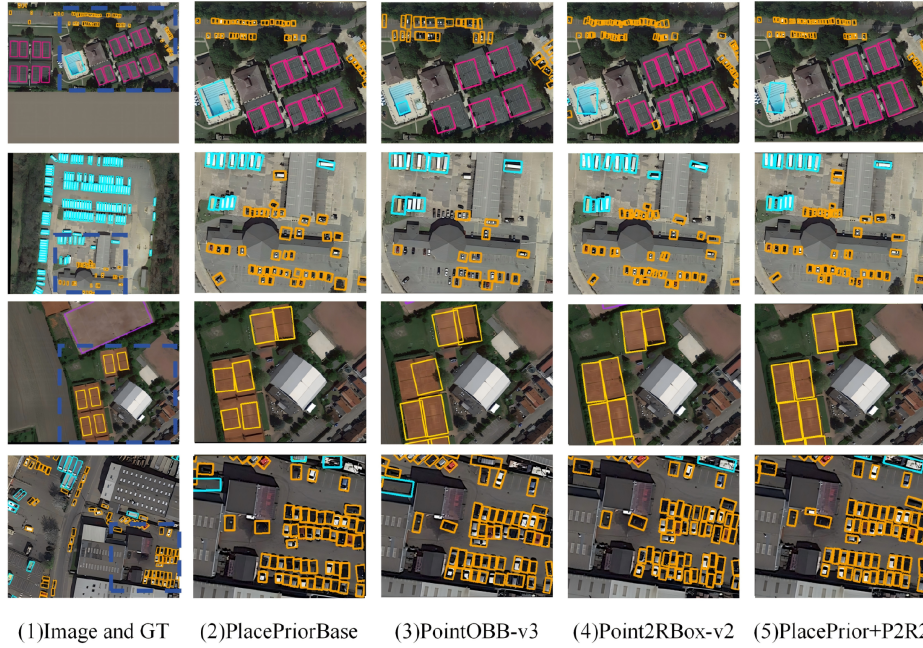


Fig. 4: The qualitative comparison against competitors on the DOTA-v1.0 dataset. The dashed boxes in col. (1) indicate enlarged regions for clearer comparisons.

Table 2: The ablation study of our loss terms on PlacePriorBase and Point2RBox-v2 [58] on a UAV dataset and a remote sensing dataset.

Modules			DroneVehicle		DOTA-v1.5	
$\mathcal{L}_{\text{RBox}}$	$\mathcal{L}_{\text{angle}}$	$\mathcal{L}_{\text{size}}$	PlacePriorBase	Point2RBox-v2	PlacePriorBase	Point2RBox-v2
			-	53.70	-	39.45
✓			38.21	54.47	30.17	40.65
✓	✓		39.53	55.54	31.69	42.36
✓		✓	39.66	55.82	31.42	42.08
✓	✓	✓	40.75	56.37	32.38	43.06

of Fig. 3 and the 4th row of Fig. 4, the RBoxes estimated by PlacePrior+P2R2 present more precise angles and sizes.

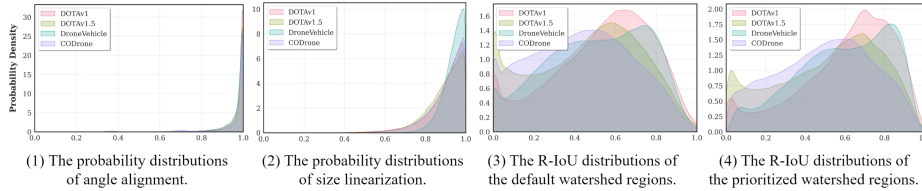
4.3 Ablation Study

To investigate the individual contributions of our loss terms, we conduct an ablation study on representative UAV and remote sensing datasets. Specifically, we conduct ablation studies on PlacePriorBase and Point2RBox-v2 [58] by adding different combinations of our loss terms.

As shown in Tab. 2, comparing Row #2 with Row #1 shows that adding our $\mathcal{L}_{\text{RBox}}$ to Point2RBox-v2 yields dramatic improvements on both datasets (*e.g.*,

Table 3: The robustness to point annotation noise perturbation.

σ	PlacePriorBase		PointOBB-v3		Point2RBox-v2		PlacePrior+P2R2	
	DOTA-v1.5	DroneVehicle	DOTA-v1.5	DroneVehicle	DOTA-v1.5	DroneVehicle	DOTA-v1.5	DroneVehicle
0%	32.38	40.75	31.25	42.43	39.45	53.70	43.06	56.37
10%	32.01	40.43	30.12	40.79	38.82	53.02	42.53	55.76
20%	31.76	40.14	28.92	38.93	38.01	52.43	41.96	54.88
30%	31.13	39.56	26.80	36.38	37.24	50.11	41.18	53.95

**Fig. 5:** The statistics on the validation of our prior knowledge across four datasets.

+1.20% for Point2RBox-v2 on DOTA-v1.5), indicating that our training-free RBox can provide beneficial supervision. Enabling either $\mathcal{L}_{\text{angle}}$ or $\mathcal{L}_{\text{size}}$ individually leads to clear performance improvements for both PlacePriorBase and Point2RBox-v2 (e.g., +1.71% by $\mathcal{L}_{\text{angle}}$ for Point2RBox-v2 on DOTA-v1.5), verifying that our loss terms provide beneficial supervisory signals based on placement prior. With all loss terms enabled, the performances are further improved, demonstrating that our loss terms are complementary to each other.

4.4 Robustness to Point Annotation Noise

We further evaluate the robustness of four point-supervised oriented object detection methods (PlacePriorBase, PointOBB-v3 [62], Point2RBox-v2 [58], and PlacePrior+P2R2) to point annotation noise on the DOTA-v1.5 [43] and DroneVehicle [38] datasets. To simulate real-world annotation jitter, coordinate perturbations are added to the ground-truth RBox centers. The random offsets are drawn from a uniform distribution $\mathcal{U}(-\sigma \cdot s, \sigma \cdot s)$, where $s = \min(w, h)$ represents the shorter side of the ground-truth RBox, and σ denotes the noise intensity. Results are summarized in Tab. 3.

Experiments show that PlacePrior+P2R2 (Point2RBox-v2 equipped with our loss terms) maintains satisfactory performance under various noise cases. As Tab. 3 shows, PlacePrior+P2R2 consistently retains the optimal overall mAP@0.5 across all evaluated noise intensities against other baselines.

4.5 In-Depth Analysis on Our Loss Terms

To further investigate the effectiveness and applicability of our loss terms, we perform statistics based on the GT RBoxes of the object instances in the training sets of four aerial datasets. Specifically, for the angle prior loss, we calculate the

Table 4: The comparison on computational efficiency in training and inference stages.

Method	Network Ops (ms)			Loss Computations (ms)				Training		Inference	
	Feat	Extra	Base	$\mathcal{L}_{base}$	$\mathcal{L}_{aux}$	$\mathcal{L}_W$	$\mathcal{L}_{ours}$	Time	Mem.	Time	Mem.
Point2RBox-v2 [58]	15.37	11.33	9.06	0.60	2.38	45.76	-	85.36	2706.7	9.22	318.2
PlacePriorBase	6.84	-	4.43	0.59	-	-	1.43	14.19	1443.7	9.21	317.9
PlacePrior+P2R2	15.34	11.29	9.07	0.60	2.34	44.23	2.44	86.24	-	9.22	318.2

Network Ops: **Feat** (Backbone & Neck); **Extra** (*Point2RBox-v2*’s heavy patches); **Base** (Augment, assignment, grouping, & head).

Loss Computations: $\mathcal{L}_{base}$ (Cls. & box loss); $\mathcal{L}_{aux}$ (Overlap, edge, & consistency losses); $\mathcal{L}_W$ (Dense Voronoi Watershed); $\mathcal{L}_{ours}$ (Our proposed efficient pseudo-label, size, and angle priors).

angular similarity between the neighborhood aligned angle and GT angle using the GT RBoxes, and plot the similarity distributions of object instances in four datasets. As shown in Fig. 5 (1), the vast majority of neighborhood aligned angles closely match the GT angles on four datasets. In particular, the value 1 on the x-axis indicates the 0° angular deviation, and the value 0.8 on the x-axis indicates the 19.6° angular deviation. For the size prior loss, we calculate the 1-D IoUs between the regressed heights and GT heights. As shown in Fig. 5 (2), the main peaks of the distributions lie in the range about (0.9, 1), indicating that most object instances generally exhibit proportional sizes.

For the RBoxes derived from watershed regions, we calculate the 2-D rotated IoUs between the derived RBoxes and GT RBoxes for both the default watershed regions and our prioritized watershed regions. As shown in Fig. 5 (3), the RBoxes derived from the default watershed regions generally achieve moderate rotated IoUs across four datasets. Compared with Fig. 5 (3), the distributions in Fig. 5 (4) dramatically shift to higher IoUs, demonstrating that our prioritized watershed regions could provide better RBoxes without training.

4.6 Efficiency Analysis

As our main focus is setting a simple baseline with readily transferred loss terms, we compare the efficiency of PlacePriorBase against the end-to-end SOTA methods in Tab. 4. Specifically, using a batch size of 1, we average various record values between the 1000-th and 1500-th training or test iterations on the same 4090 GPU to obtain stable statistics. The “Batch Time” (ms) shows the averaged time for each entire batch. Given the indispensable time overhead of the backbone network, we adopt a fine-grained batch time decomposition to explicitly reveal the efficiency bottleneck of each individual method. The “Mem.” (MB) indicates the required GPU memory.

As shown in Tab. 4, our method is highly efficient in both time and memory consumption. By providing a fine-grained batch time decomposition, we explicitly reveal that compared to Point2RBox-v2 [58], PlacePriorBase significantly accelerates the training (14.19ms *vs.* 85.36ms) by discarding heavy engineering

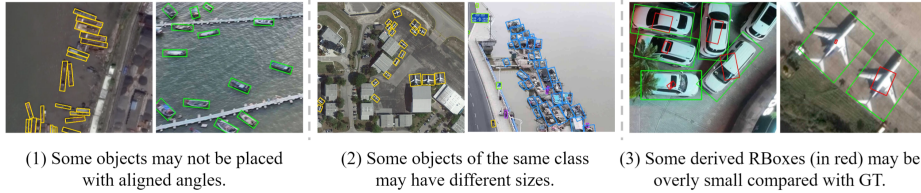


Fig. 6: The representative failure cases of our three loss terms.

patches (Extra) and complex auxiliary losses ($\mathcal{L}_W$), relying purely on our efficient geometric priors ($\mathcal{L}_{ours}$). Specifically, Point2RBox-v2 [58] incorporates image augmentation, and its time and memory consumption increase by multiple times as the number of images is multiplied.

4.7 Limitation and Discussion

Generally, our loss terms rely on prior knowledge regarding object placement in aerial images, and thus have a specific scope of application. Specifically, for the training set of the four aerial datasets, there exist cases that do not align with our assumptions. We summarize representative failure cases for three loss terms in Fig. 6. Specifically, we could see that some objects are actually not placed with aligned angles in the sub-figure (1). We may also see that the objects of the same class could have several kinds of different sizes in sub-figure (2). Besides, some RBoxes derived from watershed regions are too large or too small compared with GT RBoxes in sub-figure (3), which may be misled by the low-level textures.

5 Conclusion

In this work, we investigate point-supervised oriented object detection for aerial images. We construct PlacePriorBase, a simple yet strong baseline, by integrating three lightweight and effective loss terms with the intrinsic placement prior of aerial objects. Benefiting from these plug-and-play losses, PlacePriorBase is easy to extend and can be readily transferred to other point-supervised detectors. Extensive experiments on representative UAV and remote sensing datasets fully verify the effectiveness of PlacePriorBase and our loss terms.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grants U24A20220, 62132006, 62402201, 62461028, in part by the Natural Science Foundation of Jiangxi Province of China under Grants 20252BAC230003, 20252BAC240197, in part by the China Postdoctoral Science Foundation under Grant 2025M771495, and in part by the Early-Career Young Scientists and Technologists Project of Jiangxi Province under Grant 20244BCE52070.

References

1. Aurenhammer, F.: Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM computing surveys (CSUR)* **23**(3), 345–405 (1991)
2. Bian, Y., Ju, X., Li, J., Xu, Z., Cheng, D., Xu, Q.: Multi-patch prediction: Adapting llms for time series representation learning. In: *ICML* (2024)
3. Cao, G., Yu, X., Yu, W., Han, X., Yang, X., Li, G., Jiao, J., Han, Z.: P2rbox: Point prompt oriented object detection with sam. *arXiv preprint arXiv:2311.13128* (2023)
4. Chen, J., Luo, Z., Liu, Z., Jiang, W., Niu, L., Fang, Y.: Weak-shot keypoint estimation via keyness and correspondence transfer. In: Belgrave, D., Zhang, C., Lin, H., Pascanu, R., Koniusz, P., Ghassemi, M., Chen, N. (eds.) *Advances in Neural Information Processing Systems*. vol. 38, pp. 75279–75299. Curran Associates, Inc. (2025)
5. Chen, J., Yan, J., Fang, Y., Niu, L.: Webly supervised fine-grained classification by integrally tackling noises and subtle differences. *IEEE Transactions on Image Processing* **34**, 2641–2653 (2025)
6. Chen, P., Yu, X., Han, X., Hassan, N., Wang, K., Li, J., Zhao, J., Shi, H., Han, Z., Ye, Q.: Point-to-box network for accurate object detection via single point supervision. In: *European Conference on Computer Vision*. pp. 51–67. Springer (2022)
7. Ding, J., Xue, N., Long, Y., Xia, G.S., Lu, Q.: Learning roi transformer for oriented object detection in aerial images. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 2849–2858 (2019)
8. Feng, D., Haase-Schütz, C., Rosenbaum, L., Hertlein, H., Glaeser, C., Timm, F., Wiesbeck, W., Dietmayer, K.: Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. *IEEE Transactions on Intelligent Transportation Systems* **22**(3), 1341–1360 (2020)
9. Fu, K., Chang, Z., Zhang, Y., Xu, G., Zhang, K., Sun, X.: Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing* **161**, 294–308 (2020)
10. Gao, S., Zhang, W., Wang, Y., Guo, Q., Zhang, C., He, Y., Zhang, W.: Weakly-supervised salient object detection using point supervision. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 36, pp. 670–678 (2022)
11. Han, J., Ding, J., Li, J., Xia, G.S.: Align deep features for oriented object detection. *IEEE transactions on geoscience and remote sensing* **60**, 1–11 (2021)
12. Han, J., Ding, J., Xue, N., Xia, G.S.: 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 (2021)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
14. Hou, L., Lu, K., Yang, X., Li, Y., Xue, J.: G-rep: Gaussian representation for arbitrary-oriented object detection. *Remote Sensing* **15**(3) (2023)
15. Hu, J., Lin, J., Gong, S., Cai, W.: Relax image-specific prompt requirement in sam: A single generic prompt for segmenting camouflaged objects. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 38, pp. 12511–12518 (2024)
16. Hu, J., Lin, J., Yan, J., Gong, S.: Leveraging hallucinations to reduce manual prompt dependency in promptable segmentation. *Advances in Neural Information Processing Systems* **37**, 107171–107197 (2024)

17. Iqbal, J., Munir, M.A., Mahmood, A., Ali, A.R., Ali, M.: Leveraging orientation for weakly supervised object detection with application to firearm localization. *Neurocomputing* **440**, 310–320 (2021)
18. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics yolov5 (2020), <https://github.com/ultralytics/yolov5>, yOLOv5s model variant. Accessed: 2026-06-27
19. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
20. Li, Q., He, X., Shu, X., Yu, Y., Chen, D., Chen, Y., Yang, X.: A simple aerial detection baseline of multimodal language models. In: IGARSS 2025-2025 IEEE International Geoscience and Remote Sensing Symposium. pp. 6833–6837. IEEE (2025)
21. Li, W., Chen, Y., Hu, K., Zhu, J.: Oriented reppoints for aerial object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1829–1838 (2022)
22. Li, W., Yuan, Y., Wang, S., Zhu, J., Li, J., Liu, J., Zhang, L.: Point2mask: Point-supervised panoptic segmentation via optimal transport. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 572–581 (2023)
23. Liao, Z., Li, J., Niu, L., Zhang, L.: Align and aggregate: Compositional reasoning with video alignment and answer aggregation for video question-answering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13395–13404 (2024)
24. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)
25. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**(2), 318–327 (2020)
26. Liu, M., Zhang, P., Liu, Y., Zhang, W., Zhou, Y., Liao, N., Gong, Z., Luo, J., Wang, Z., Yu, Y., Yang, X.: Partial weakly-supervised oriented object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 27644–27654 (June 2026)
27. Liu, N., Xu, X., Su, Y., Zhang, H., Li, H.C.: Pointsam: Pointly-supervised segment anything model for remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* **63**, 1–15 (2025)
28. Liu, Z., Yuan, L., Weng, L., Yang, Y.: A high resolution optical satellite image dataset for ship recognition and some new baselines. In: International conference on pattern recognition applications and methods. vol. 2, pp. 324–331. SciTePress (2017)
29. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
30. Lou, Z., Zhang, R., Luo, Z., Cao, Q., Qian, F., Chen, J., Fang, Y.: Infrared-privileged uav detection via cross-modal vector-quantization. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 7521–7529 (2026)
31. Luo, J., Yang, X., Yu, Y., Li, Q., Yan, J., Li, Y.: Pointobb: Learning oriented object detection via single point supervision. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16730–16740 (2024)
32. Lyu, C., Zhang, W., Huang, H., Zhou, Y., Wang, Y., Liu, Y., Zhang, S., Chen, K.: Rtmddet: An empirical study of designing real-time object detectors. arXiv preprint arXiv:2212.07784 **7** (2022)

33. Murrugarra-Llerena, J., Kirsten, L.N., Zeni, L.F., Jung, C.R.: Probabilistic intersection-over-union for training and evaluation of oriented object detectors. *IEEE Transactions on Image Processing* **33**, 671–681 (2024)
34. Pan, X., Ren, Y., Sheng, K., Dong, W., Yuan, H., Guo, X., Ma, C., Xu, C.: Dynamic refinement network for oriented and densely packed object detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 11207–11216 (2020)
35. Qian, W., Yang, X., Peng, S., Yan, J., Guo, Y.: Learning modulated loss for rotated object detection. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 35, pp. 2458–2466 (2021)
36. Ren, B., Yang, X., Yu, Y., Luo, J., Deng, Z.: Pointobb-v2: Towards simpler, faster, and stronger single point supervised oriented object detection. In: Yue, Y., Garg, A., Peng, N., Sha, F., Yu, R. (eds.) *International Conference on Learning Representations*. vol. 2025, pp. 39843–39857 (2025)
37. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* **28** (2015)
38. Sun, Y., Cao, B., Zhu, P., Hu, Q.: Drone-based rgb-infrared cross-modality vehicle detection via uncertainty-aware learning. *IEEE Transactions on Circuits and Systems for Video Technology* **32**(10), 6700–6713 (2022)
39. Sun, Y., Ran, J., Yang, F., Gao, C., Kurozumi, T., Kimata, H., Ye, Z.: Oriented object detection for remote sensing images based on weakly supervised learning. In: *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. pp. 1–6. IEEE (2021)
40. Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 9627–9636 (2019)
41. Vincent, L., Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **13**(06), 583–598 (1991)
42. Wang, L., Zhan, Y., Lin, X., Yu, B., Ding, L., Zhu, J., Tao, D.: Explicit and implicit box equivariance learning for weakly-supervised rotated object detection. *IEEE Transactions on Emerging Topics in Computational Intelligence* **9**(1), 509–521 (2024)
43. Xia, G.S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L.: Dots: 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)
44. Xie, X., Cheng, G., Wang, J., Yao, X., Han, J.: Oriented r-cnn for object detection. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 3520–3529 (2021)
45. Yang, X., Hou, L., Zhou, Y., Wang, W., Yan, J.: Dense label encoding for boundary discontinuity free rotation detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 15819–15829 (2021)
46. Yang, X., Yan, J.: Arbitrary-oriented object detection with circular smooth label. In: *European conference on computer vision*. pp. 677–694. Springer (2020)
47. Yang, X., Yan, J.: On the arbitrary-oriented object detection: Classification based approaches revisited. *International Journal of Computer Vision* **130**(5), 1340–1365 (2022)

48. Yang, X., Yan, J., Liao, W., Yang, X., Tang, J., He, T.: Scrdet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing. *IEEE transactions on pattern analysis and machine intelligence* **45**(2), 2384–2399 (2022)
49. Yang, X., Yan, J., Ming, Q., Wang, W., Zhang, X., Tian, Q.: Rethinking rotated object detection with gaussian wasserstein distance loss. In: *International conference on machine learning*. pp. 11830–11841. PMLR (2021)
50. Yang, X., Yang, J., Yan, J., Zhang, Y., Zhang, T., Guo, Z., Sun, X., Fu, K.: Scrdet: Towards more robust detection for small, cluttered and rotated objects. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 8232–8241 (2019)
51. Yang, X., Yang, X., Yang, J., Ming, Q., Wang, W., Tian, Q., Yan, J.: Learning high-precision bounding box for rotated object detection via kullback-leibler divergence. *Advances in Neural Information Processing Systems* **34**, 18381–18394 (2021)
52. Yang, X., Zhang, G., Li, W., Wang, X., Zhou, Y., Yan, J.: H2rbox: Horizontal box annotation is all you need for oriented object detection. *arXiv preprint arXiv:2210.06742* (2022)
53. Yang, X., Zhang, G., Yang, X., Zhou, Y., Wang, W., Tang, J., He, T., Yan, J.: Detecting rotated objects as gaussian distributions and its 3-d generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(4), 4335–4354 (2023)
54. Yang, X., Zhou, Y., Zhang, G., Yang, J., Wang, W., Yan, J., Zhang, X., Tian, Q.: The kfio loss for rotated object detection. *arXiv preprint arXiv:2201.12558* (2022)
55. Yang, Z., Liu, S., Hu, H., Wang, L., Lin, S.: Reppoints: Point set representation for object detection. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 9657–9666 (2019)
56. Ye, K., Tang, H., Liu, B., Dai, P., Cao, L., Ji, R.: More clear, more flexible, more precise: a comprehensive oriented object detection benchmark for uav. *arXiv preprint arXiv:2504.20032* (2025)
57. Yu, Y., Da, F.: 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)
58. Yu, Y., Ren, B., Zhang, P., Liu, M., Luo, J., Zhang, S., Da, F., Yan, J., Yang, X.: 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 (2025)
59. Yu, Y., Yang, X., Li, Q., Da, F., Dai, J., Qiao, Y., Yan, J.: Point2rbox: Combine knowledge from synthetic visual patterns for end-to-end oriented object detection with single point supervision. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 16783–16793 (2024)
60. Yu, Y., Yang, X., Li, Q., Zhou, Y., Da, F., Yan, J.: H2rbox-v2: Incorporating symmetry for boosting horizontal box supervised oriented object detection. *Advances in Neural Information Processing Systems* **36**, 59137–59150 (2023)
61. Yu, Y., Yang, X., Li, Y., Han, Z., Da, F., Yan, J.: Wholly-wood: Wholly leveraging diversified-quality labels for weakly-supervised oriented object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **47**(6), 4438–4454 (2025)
62. Zhang, P., Luo, J., Yang, X., Yu, Y., Li, Q., Zhou, Y., Jia, X., Lu, X., Chen, J., Li, X., et al.: Pointobb-v3: Expanding performance boundaries of single point-supervised oriented object detection. *International Journal of Computer Vision* **133**(9), 6108–6128 (2025)

63. Zhang, S., Long, J., Xu, Y., Mei, S.: Pmho: Point-supervised oriented object detection based on segmentation-driven proposal generation. *IEEE Transactions on Geoscience and Remote Sensing* **62**, 1–18 (2024)
64. Zhang, T., Fan, Z., Liu, M., Zhang, X., Lu, X., Li, W., Zhou, Y., Yu, Y., Li, X., Yan, J., et al.: Point2rbox-v3: Self-bootstrapping from point annotations via integrated pseudo-label refinement and utilization. *arXiv preprint arXiv:2509.26281* (2025)
65. Zhang, W., Liu, X., Liu, N., Liu, M., Liao, W., Xu, C., Yang, X.: SPWOOD: Sparse partial weakly-supervised oriented object detection. In: *The Fourteenth International Conference on Learning Representations* (2026)
66. Zhu, T., Ferenczi, B., Purkait, P., Drummond, T., Rezatofghi, H., Van Den Hengel, A.: Knowledge combination to learn rotated detection without rotated annotation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 15518–15527 (2023)