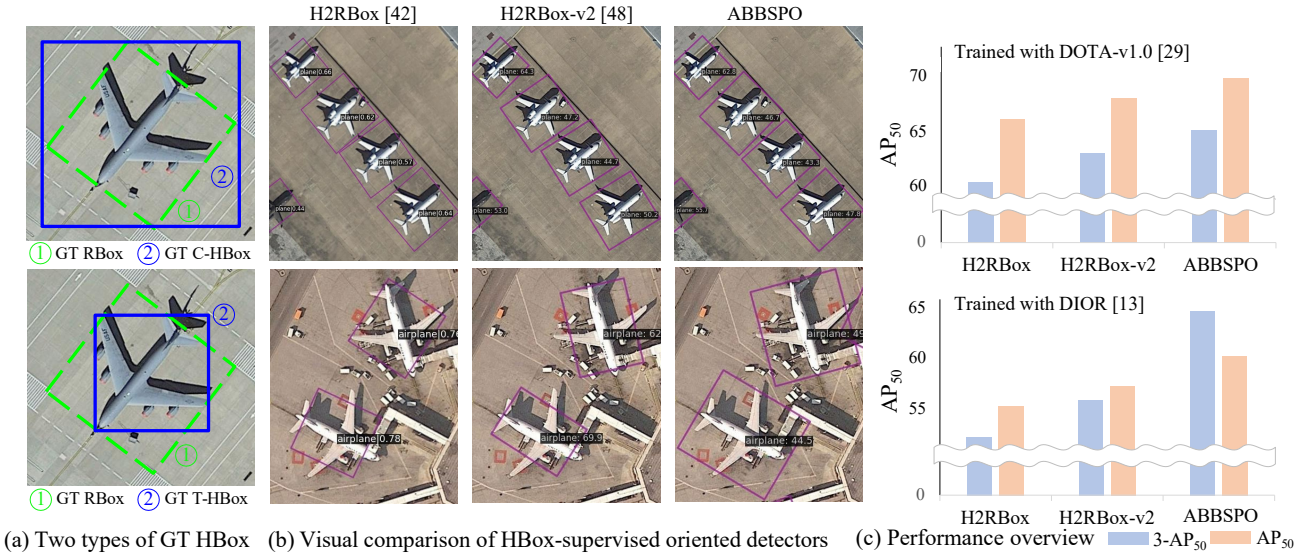
ABBSPO: Adaptive Bounding Box Scaling and Symmetric Prior based Orientation Prediction for Detecting Aerial Image Objects

Woojin Lee[1]* Hyugjae Chang[1]* Jaeho Moon[1] Jaehyup Lee[2]† Munchurl Kim[1]†
[1]KAIST [2]KNU
{woojin412, hmnc97, jaeho.moon, mkimee}@kaist.ac.kr jaehyuplee@knu.ac.kr
https://kaist-viclab.github.io/ABBSPO_site/



(a) Two types of GT HBox (b) Visual comparison of HBox-supervised oriented detectors (c) Performance overview

Figure 1. Performance comparison of HBox-supervised orientated detectors. (a) Top: A coarse horizontal bounding box (C-HBox) (②) and its corresponding rotated bounding box (RBox) (①). Bottom: A tight horizontal bounding box (T-HBox)(②) and its corresponding RBox (①). (b) Our ABBSPO is capable of accurately detecting both orientations and scales for GT C-HBoxes and T-HBoxes. (c) Average Precision (AP_{50}) for H2RBox [42], H2RBox-v2 [48], and our ABBSPO. 3-AP_{50} represents the mean AP_{50} for three complex shaped objects: (i) DIOR: 'airplane', 'expressway service area', and 'overpass' and (ii) DOTA: 'plane', 'swimming pool', and 'helicopter'.

Abstract

Weakly supervised Oriented Object Detection (WS-OOD) has gained attention as a cost-effective alternative to fully supervised methods, providing efficiency and high accuracy. Among weakly supervised approaches, horizontal bounding box (HBox) supervised OOD stands out for its ability to directly leverage existing HBox annotations while achieving the highest accuracy under weak supervision settings. This paper introduces adaptive bounding box scaling and symmetry-prior-based orientation prediction, called ABBSPO that is a framework for WS-OOD. Our ABBSPO addresses the limitations of previous HBox-supervised OOD methods, which compare ground truth (GT) HBoxes directly with predicted RBoxes' minimum circumscribed rectangles, often leading to inaccuracies. To overcome this, we propose: (i) Adaptive Bounding Box Scaling (ABBS) that appropriately scales the GT HBoxes to optimize for the size of each predicted RBox, ensuring more accurate prediction for RBoxes' scales; and (ii) a Symmetric Prior Angle (SPA) loss that uses the inherent symmetry of aerial objects for self-supervised learning, addressing the issue in previous methods where learning fails if they consistently make incorrect predictions for all three augmented views (original, rotated, and flipped). Extensive experimental results demonstrate that our ABBSPO achieves state-of-the-art results, outperforming existing methods.

*Co-first authors (equal contribution).
†Co-corresponding authors.
[1]Korea Advanced Institute of Science and Technology (KAIST).
[2]Kyungpook National University (KNU).

1. Introduction

Object detection often leverages supervised learning with ground truth horizontal bounding box labels (GT HBoxes) to locate the objects of interest. However, the usage of GT HBoxes limits the precise localization of the objects with their orientations and tight surrounding boundaries, especially for objects such as airplanes and ships of various orientations in aerial images. To handle object detection as an oriented object detection problem, more precise rotated bounding box labels (GT RBoxes) are required, which is very costly to generate [49]. So, to mitigate this challenge, previous methods [17, 23, 42, 47–49] have explored weakly supervised oriented object detection (OOD) that utilizes less expensive forms of annotations, such as image-level, point and HBox annotations. Among these, the use of HBoxes is the most popular due to their widespread availability in existing public datasets [4, 9, 12, 21, 30, 31] to predict the RBoxes for objects of interest. So, this approach can detour the costly process of generating GT RBoxes.

The previous weakly supervised (WS) learning of OOD [22, 27, 42, 48] utilizes GT HBoxes in the forms of *coarse* HBoxes, called GT C-HBoxes, as supervision to compare with the HBoxes derived as the minimum circumscribed rectangles from the predicted RBoxes by their OOD models. As shown in the upper figure of Fig. 1-(a), the GT C-HBoxes are defined as *coarse* horizontal bounding boxes that loosely encompass the boundaries of objects (not tightly bounded). The GT HBoxes of the DOTA [29] dataset are in the forms of C-HBoxes which are derived as the minimum circumscribed horizontal bounding boxes of their GT RBoxes. However, when the previous OOD methods [42, 48] are supervised with the other GT HBoxes that are in the form of *tight* HBoxes, called GT T-HBoxes (e.g. DIOR dataset [13]), as shown in the bottom figure of Fig. 1-(a), we found that their performances are significantly degraded because GT T-HBoxes tend to have different scales, compared to those of GT C-HBoxes (see Fig. 1-(c)). As shown in Fig. 1-(b), this causes the previous methods to predict either RBoxes with accurate orientations but inaccurate scales smaller than the sizes of their corresponding objects, or the RBoxes with inaccurate (close to horizontal) orientations but somewhat accurate scales (almost the same as HBoxes).

To overcome the above limitations of the previous WS-OOD methods, we propose an adaptive bounding box scaling and symmetry-prior-based orientation prediction, called as ABBSPO, as a WS-OOD framework that can be effectively trained with either GT C-HBoxes or GT T-Hboxes for aerial images. For this, (i) a novel Adaptive Bounding Box Scaling (ABBS) module is designed to have the flexibility of adjusting the GT HBoxes for each object into random sizes and then selecting the optimal scaled GT HBoxes that allow it to encompass the predicted RBoxes. Note that the previous methods are not possible to have such flexibility for the adjustment of GT HBoxes; (ii) An angle learning module is proposed in a self-supervised manner that utilizes the symmetric priors of the objects that open appear in top-down views of aerial images. As shown in Figs 1-(b) and (c), Our proposed method predicts accurate orientation and surrounding boxes of objects for both cases of using GT C-HBoxes and GT T-HBoxes, outperforming the previous methods in angle accuracy and localization in terms of average precision (AP). Our contributions are summarized as:

- To the best of our knowledge, our work is the *first* to address the limitations of previous weakly supervised OOD learning methods with T-HBoxes as GT. To overcome this, we propose a novel weakly supervised OOD method that can be effectively trained with T-HBoxes or C-HBoxes that can be cheaply annotated as GT;
- The adaptive bounding box scaling (ABBS) module is proposed to flexibly adjust the HBox (GT) for each object toward an appropriately scaled HBox. This allows part of the predicted RBoxes to place outside the T-HBox (GT), yielding precise RBox prediction;
- A symmetric prior angle (SPA) loss is presented to enhance the orientation prediction accuracy by leveraging the symmetric priors of the objects in aerial images;
- Our method significantly outperforms the state-of-the-art OOD methods using weakly supervised learning with HBoxes (GT) for aerial datasets.

2. Related Work

2.1. RBox-supervised Oriented Object Detection

Oriented Object Detection (OOD) has gained significant attention, leading to extensive research in RBox-supervised methods (using GT RBoxes) such as Rotated RetinaNet [20], Rotated FCOS [24], R^3Det [37], ROI Transformer [6], ReDet [8], and S^2A-Net [7]. Rotated FCOS [24] improves OOD performance by introducing center-ness, which assigns weights to samples based on their proposal locations, thereby emphasizing well-positioned proposals. Oriented-RepPoints methods [3, 14, 43], in contrast, utilize flexible receptive fields to extract key object points. However, a common challenge in RBox-supervised OOD methods is the boundary discontinuity problem that arises from the definition and prediction of angle parameters (θ) [33, 34]. To address this, several methods modified the ways of defining the RBox representations, such as Gaussian distributions [25, 26, 35, 36, 38–41], thereby avoiding straightforward regression of angle parameters. On the other hand, in weakly supervised learning, the boundary discontinuity issue does not arise thanks to the absence of direct RBox supervision, allowing for more stable angle predictions without the need for complex mitigation strategies.

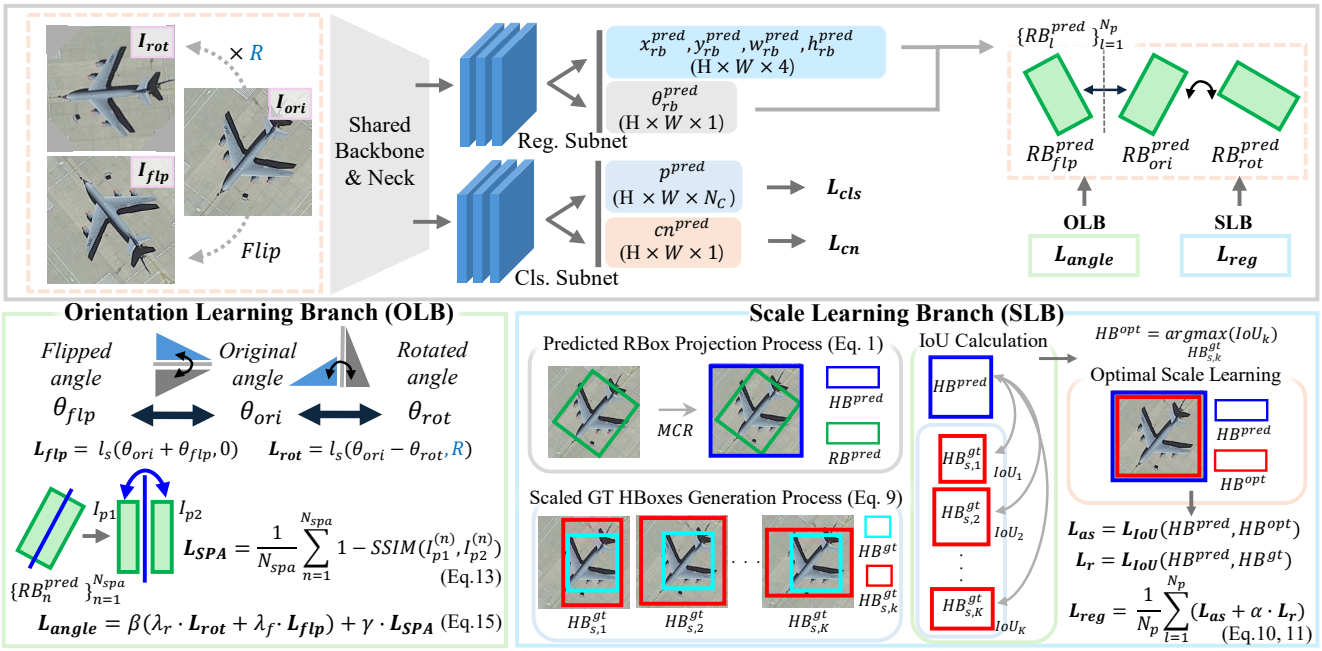


Figure 2. **Overall pipeline of our ABBSPO framework.** Our ABBSPO leverages weakly supervised learning from HBox annotations to accurately predict RBoxes. The framework incorporates the Orientation Learning Branch (OLB) for precise angle estimation, using the Symmetric Prior Angle (SPA) loss, and the Scale Learning Branch (SLB) for optimal scale adjustment via the Adaptive Bounding Box Scaling (ABBS) module. The framework supports both C-HBox and T-HBox ground truths, ensuring robust and accurate predictions.

2.2. Weakly-supervised Orientd Object Detection

Weakly supervised OOD methods learn to predict RBoxes without directly utilizing GT RBoxes. The approaches in this domain are primarily categorized based on the types of labels they employ: image-based [23], point-based [17, 47], and HBox-based [42, 48].

Image-based supervision. WSODet [23], aims to generate pseudo-RBoxes without explicit localization supervision, thus encountering significant limitations when relying solely on image labels, especially for the scenes with numerous and diverse object types.

Point-based supervision. By leveraging one representative point at the center location as a label for each object, point label-based methods offer the advantage of being cost-effective [1, 2, 11, 16, 19, 44]. PointOBB [17] estimates angles from geometric relationships across original, rotated, and flipped views, and determines scales by analyzing proposal distributions between original and scaled input images. PointOBB-v2 [18] improves single point supervision by refining pseudo-label generation, leading to enhanced efficiency and accuracy. Point2RBox [47] employed fundamental patterns as priors to guide the regression of RBoxes. Point-based methods are cost-effective and straightforward, but still struggle with limited supervision.

HBox-based supervision. As annotating HBoxes is more straightforward than RBoxes, the HBox-supervised OOD has gained increasing attention in recent studies. H2RBox [42] utilized rotated views from the original view and provided self-supervision of object orientations without requiring the GT angles. H2RBox-v2 [48] expanded the use of geometric relationships between views by adding flipped views. These methods learn to predict RBoxes by converting minimum circumscribed HBoxes that encompass the predicted RBoxes to directly compare IoU with the GT HBoxes, thereby enabling HBox-supervised OOD. However, these methods only guarantee performance when trained with GT C-HBoxes that are derived from GT RBoxes. These methods struggle to learn precise OOD when being trained with GT T-HBoxes, because of the significant gap between the GT T-HBoxes and the HBoxes derived from predicted RBoxes. To address the issue, we propose an ABBS module to effectively handle both types of GTs (C-HBoxes and T-HBoxes).

3. Method

3.1. Overall Pipeline

Weakly supervised OOD aims to predict RBoxes using less expensive annotations such as HBoxes. Existing methods such as H2RBox [42] and its improved version H2RBox-v2 [48] have laid the foundation for directly predicting RBoxes from HBoxes. Our proposed pipeline builds upon the H2RBox-v2 [48] framework to effectively enable weakly

supervised OOD from either C-HBoxes or T-Boxes. Figure 2 depicts the conceptual framework of our weakly supervised OOD method. Given an input image I_{ori} and its rotated and flipped version I_{rot} and I_{flp}, our proposed pipeline obtains the RBox for each input view (I_{ori}, I_{rot}, I_{flp}), including the center position (x,y), size (w,h), angle (θ), class scores (p), and the center-ness (cn). To classify each detected object, we follow FCOS [24] by supervising both the classification (p) and the center-ness (cn). The angle (θ) prediction is obtained by using the method proposed in PSC [46]. Our contribution mainly lies in the supervision for localization, consisting of two branches: a scale learning branch (SLB) and an orientation learning branch (OLB).

In the SLB, the adaptive bounding box scaling (ABBS) module addresses the relationships between GT HBoxes and predicted RBoxes. This ABBS module provides proper minimum circumscribed rectangle for the accurately predicted RBoxes, by adaptively scaling the HBoxes based on a predefined scale range. The OLB guides accurate prediction of object orientation by utilizing three input views (I_{ori}, I_{rot}, I_{flp}), following H2RBox-v2 [48]. Additionally, the OLB utilizes these orientation predictions for our symmetric prior angle (SPA) loss, which leverages the inherent left-right symmetry of objects in aerial images. The SPA loss enforces to further adjust the orientations of the predicted RBoxes to be aligned with the orientations of the symmetric objects such as airplanes, ships, ground track fields etc.

3.2. Adaptive Bounding Box Scaling Module

In Fig. 2, 'Scale Learning Branch' illustrates the conceptual process of our adaptive bounding box scaling module (ABBS) module. In HBox-supervised OOD learning, the predicted RBoxes (RB^{pred}) must be compared with the GT HBoxes (HB^{gt}). Since they cannot be directly compared, RB^{pred} is first converted to HB^{pred}, defined as the minimum circumscribed HBox of RB^{pred} as:

$$HB^{\text{pred}} = MCR(RB^{\text{pred}}), \tag{1}$$

where $MCR(\cdot)$ is an operator converting RB to the minimum circumscribed HB, allowing HB^{pred} to be compared with HB^{gt}. HB^{pred} and RB^{pred} are given by:

$$\begin{aligned} RB^{\text{pred}} &= [x_{rb}^{\text{pred}}, y_{rb}^{\text{pred}}, w_{rb}^{\text{pred}}, h_{rb}^{\text{pred}}, \theta_{rb}^{\text{pred}}], \\ HB^{\text{pred}} &= [x_{hb}^{\text{pred}}, y_{hb}^{\text{pred}}, w_{hb}^{\text{pred}}, h_{hb}^{\text{pred}}], \end{aligned} \tag{2}$$

where $(x_{rb}^{\text{pred}}, y_{rb}^{\text{pred}})$ and $(x_{hb}^{\text{pred}}, y_{hb}^{\text{pred}})$ are the centers of RB^{pred} and HB^{pred}, respectively. The width w and height h of HB^{pred} can be computed as:

$$\begin{aligned} w_{hb}^{\text{pred}} &= w_{rb}^{\text{pred}} |\cos\theta_{rb}^{\text{pred}}| + h_{rb}^{\text{pred}} |\sin\theta_{rb}^{\text{pred}}|, \\ h_{hb}^{\text{pred}} &= w_{rb}^{\text{pred}} |\sin\theta_{rb}^{\text{pred}}| + h_{rb}^{\text{pred}} |\cos\theta_{rb}^{\text{pred}}|. \end{aligned} \tag{3}$$

If RB^{opt} is defined as the tightly surrounding object boundary RBox with the precise orientation, then we have

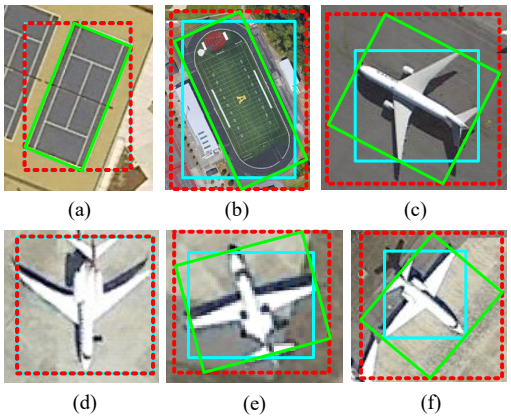


Figure 3. Analysis of scale adjustment function ($f(\cdot)$) based on the shape and angle of objects. (a) rectangular shape, (b) rounded rectangular shape, (c) complex shape, (d) horizontal orientation, (e) slightly tilted orientation, (f) diagonal orientation. The cyan solid box, green solid box and red dotted box represent GT T-HBox, RBox and adjusted GT HBox, respectively.

$HB^{\text{opt}} = MCR(RB^{\text{opt}})$ for the 'Predicted RBox Projection Process' block in the SLB of Fig. 2. When the size of HB^{gt} is larger or smaller than that of HB^{opt}, the model needs to adaptively adjust and find the optimal scale within a predefined range of scale variations. We propose an ABBS module that estimates RB^{opt} by adaptively adjusting the scale of HB^{gt}. Notably, even if RB^{pred} is accurately estimated, its HB^{pred} may not overlap well with HB^{gt}, leading to a low Intersection over Union (IoU) value. Enforcing HB^{pred} to match HB^{gt} can cause a misalignment with RB^{pred} because HB^{gt} may not be ideal in estimating RB^{opt}. To address this, our ABBS module adaptively scales and adjusts HB^{gt} in the context of RB^{pred}, rather than forcing HB^{pred} to match HB^{gt}.

For the detailed explanation of our ABBS module, we first define a set of scaled versions of HB^{gt} for the 'Scaled GT HBoxes Generation Process' in the SLB of Fig. 2 as:

$$\mathbf{HB}_{\text{s}}^{\text{gt}} = \{HB_{\text{s},1}^{\text{gt}}, HB_{\text{s},2}^{\text{gt}}, \cdots, HB_{\text{s,K}}^{\text{gt}}\}, \tag{4}$$

where $HB_{\text{s},k}^{\text{gt}}$ is the k-th scaled version of HB^{gt} and K is the total number of scaled variations of HB^{gt}. $\mathbf{HB}_{\text{S}}^{\text{gt}}$ is determined as the combinations of *angle-adjusted* width and height scale factors, $\{s_{adj,i}^{w}\}_{i=1}^{N_s}$ and $\{s_{adj,j}^{h}\}_{j=1}^{N_s}$, that are transformed from basic width and height scale factors,$\{s_i^{w}\}_{i=1}^{N_s}$ and $\{s_j^{h}\}_{j=1}^{N_s}$ by considering angle prediction. Basic scale factors are uniformly spaced in a predefined scale range as:

$$S_w = \{s_1^{w}, s_2^{w}, \cdots, s_{N_s}^{w}\},\ S_h = \{s_1^{h}, s_2^{h}, \cdots, s_{N_s}^{h}\}, \tag{5}$$

where S_w and S_h are the sets of basic width and height scale factors, respectively. s_i^{w} and s_i^{h} are calculated as:

$$\begin{aligned} s_i^{w} &= s_1^{w} + (s_{N_s}^{w} - s_1^{w})/(N_s - 1) \cdot (i - 1), \\ s_j^{h} &= s_1^{h} + (s_{N_s}^{h} - s_1^{h})/(N_s - 1) \cdot (j - 1), \end{aligned} \tag{6}$$

where $s^w_{N_s} = s^h_{N_s}$ is the predefined largest basic scale factor for both width and height of HB^{gt}, and N_s is the number of uniform quantization for the both range $[s^w_1..s^w_{N_s}]$ and $[s^h_1..s^h_{N_s}]$. In order to generate $HB^{\text{gt}}_{s,i}$, we transform the basic width and height scale factors, $\{s^w_i\}^{N_s}_{i=1}$ and $\{s^h_j\}^{N_s}_{j=1}$, into *angle-adjusted* width and height scale factors, $\{s^w_{adj,i}\}^{N_s}_{i=1}$ and $\{s^h_{adj,j}\}^{N_s}_{j=1}$, using the predicted angle θ^{pred} through the scale adjustment function f:

$$s^w_{adj,i} = f(\theta^{\text{pred}}_{rb}, s^w_i),\ s^h_{adj,j} = f(\theta^{\text{pred}}_{rb}, s^h_j). \tag{7}$$

To define $f(\cdot)$, it's essential to consider the object types and rotation angles. Fig. 3 shows the effect of scale adjustments on T-HBoxes for three object types: (i) For rectangular objects like tennis courts (Fig. 3-(a)), the adjusted T-HBox (red dotted box) aligns precisely with the GT T-HBox (cyan solid box) and tightly circumscribes the optimal RBox (green solid box); (ii) For rounded rectangular objects (Fig. 3-(b)), the optimal RBox slightly exceeds the GT T-HBox; (iii) Complex shapes like airplanes (Fig. 3-(c)) show a larger discrepancy, with parts of the optimal RBox lying outside the GT T-HBox. Furthermore, scale adjustments also depend on rotation angles: (i) Fig. 3-(d) for a vertically (or horizontally) aligned airplane, the GT T-HBox and optimal RBox are identical; (ii) For Fig. 3-(e) with a small rotation angle, they differ slightly; (iii) For Fig. 3-(f) with a larger angle, the difference is more pronounced. Therefore, to take the object's shape types and orientation degrees into account for the scale adjustment for the widths and heights of T-HBoxes, $f(\cdot)$ in Eq. 7 is defined as:

$$f(\theta, s) = \begin{cases} \frac{4}{\pi}(s-1)\cdot\theta + 1, & \text{if } 0 \le \theta < \frac{\pi}{4}, \\ \frac{4}{\pi}(1-s)\cdot\theta + (2s-1), & \text{if } \frac{\pi}{4} \le \theta < \frac{\pi}{2}, \end{cases} \tag{8}$$

where the angle range is set to $\theta \in [0, \pi/2)$ due to the periodicity of the angle. According to $f(\cdot)$ and Eq. 7, $HB^{\text{gt}}_{s,k}$ in $\mathbf{HB}^{\text{gt}}_s$ can be expressed as:

$$HB^{\text{gt}}_{s,k} = [x^{\text{gt}}, y^{\text{gt}}, w^{\text{gt}} \cdot s^w_{adj,i},\ h^{\text{gt}} \cdot s^h_{adj,j}], \tag{9}$$

where $(x^{\text{gt}}, y^{\text{gt}})$ is the center point, and w^{gt} and h^{gt} are width and height of HB^{gt}. As shown in the 'IoU Calculation' and 'Optimal Scale Learning' blocks in the SLB of Fig. 2, HB^{opt} among $\{HB^{\text{gt}}_{s,k}\}^K_{k=1}$ can be determined which minimizes the IoU loss for all proposals by an ABBS loss as:

$$\mathcal{L}_{\text{as}} = \frac{1}{N_p}\sum_{l=1}^{N_p} \min_{\substack{s^w_i \in S_w, \\ s^h_j \in S_h}} \mathcal{L}_{\text{IoU}}\left(HB^{\text{pred}}_l, HB^{\text{gt},l}_{s,k}(s^w_i, s^h_j)\right), \tag{10}$$

where N_p is the total number of proposals for input $\mathbf{I}$. HB^{pred}_l is HB^{pred} for l-th proposal, and $HB^{\text{gt},l}_{s,k}(s^w_i, s^h_j)$ is k-th scaled HB^{gt}, as $HB^{\text{gt}}_{s,k}$, whose width and height are scaled for l-th proposal according to Eq. 7 to Eq. 9. Finally, by adding a regularization term using the IoU loss between HB^{pred} and non-scaled HB^{gt}, we formed the regression loss as:

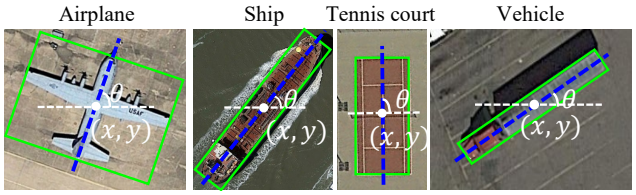


Figure 4. **Examples of symmetric objects in aerial images.** In SPA loss, the *x,y* coordinates and angle θ are used to define the symmetry axis, splitting the object into two parts for comparison.

$$\mathcal{L}_{\text{reg}} = \mathcal{L}_{\text{as}} + \alpha \cdot (1/N_p)\sum_{l=1}^{N_p} \mathcal{L}_{\text{IoU}}\left(HB^{\text{pred}}_l, HB^{\text{gt},l}\right), \tag{11}$$

where α is a hyperparameter which is set to 0.01 by default.

3.3. Symmetric Prior Angle Loss

In aerial images, objects such as airplanes, ships, tennis courts, and vehicles are often captured from top-down viewpoints, where most of these objects exhibit symmetries in their appearance, as shown in Fig. 4.

In the previous pipelines [42, 48], both the regression loss associated with the bounding box's center point, width, and height, and the angle loss for accurate angle prediction were trained in a balanced manner. However, they tended to inaccurately predict the angles by maximizing the bounding box's IoUs at the same time. This issue stems from the fact that, since the angles could not be directly supervised due to the absence of angle annotations, the angles were indirectly supervised from augmented views with rotations and flips. This is problematic because, when the difference between two predicted angles for the same object in the original view and its rotated view are equal to the rotation angle applied for the original view, the angle loss is zero although the predicted angles are inaccurate.

To mitigate such a predicted angle ambiguity, we propose a symmetric prior angle (SPA) loss. Based on the SPA loss, the model can be trained to predict precise angles by *indirectly* utilizing the object's symmetric characteristics. As shown in Fig. 4, the detected objects are symmetric against the symmetry axes (blue-dotted lines) passing the center points of their RBoxes. That is, the pixel contents in the two parts divided by the symmetry axis of the RBox are compared in similarity whose difference is used as supervision for our SPA loss. It is noted that our SPA loss utilizes only symmetric objects, incorporating the symmetry prior from GT class labels for proposals identified as symmetric, such as 'airplane,' 'ship,' 'vehicle,' and 'tennis court'

To avoid applying the SPA loss when RB^{pred} are inaccurate for the respective objects, we first check the fidelity scores of proposal, and *sample* the Top-k proposals as supervision in the SPA loss as:

$$\{RB_n^{pred}\}_{n=1}^{N_{\text{spa}}} = \text{Top-}k\left(\{RB_l^{pred}\}_{l=1}^{N_p} \mid \text{sc}_{\text{cls}}^i + \text{sc}_{\text{loc}}^i\right) \quad (12)$$

where $\text{sc}_{\text{cls}}^{(i)}$ and $\text{sc}_{\text{loc}}^{(i)}$ are the classification and localization scores for the l-th RB^{pred}, and N_p is the total number of proposals. From Eq. 12, the selected N_{spa} proposals are considered in the SPA loss by which the predicted angles of RB^{pred} (θ_{rb}^{pred}) are enforced to align with the orientations of objects in the sense of maximizing the similarity, Structural Similarity Index (SSIM [28]), between the pixel contents in the two parts of each RB^{pred}. It should be noted that, even in cases where symmetric objects may not appear perfectly symmetric due to contextual factors like shadows or asymmetrical cargo arrangements, their symmetry is still maintained by the inherent structural symmetry between the two parts. Our SPA loss is defined as:

$$\mathcal{L}_{\text{SPA}} = (1/N_{\text{spa}}) \sum_{n=1}^{N_{\text{spa}}} \left(1 - \text{SSIM}(I_{p1}^{(n)}, I_{p2}^{(n)})\right) \quad (13)$$

To remove the influence of object sizes in L_{SPA} computation, the proposals (RB^{pred}) are projected onto a fixed-size grid of 50×50. Then, the pixel content (I_{p1}) in one part of the proposal's projection is compared with that (I_{p2}) of the other part that is flipped before the comparison.

3.4. Loss Functions

In the orientation learning branch (OLB), two angle-based losses [48], $\mathcal{L}_{\text{rot}}$ and $\mathcal{L}_{\text{flp}}$, are adopted to leverage the consistency between the original, rotated, and flipped views of each object proposal. For the rotated and flipped views, $\mathcal{L}_{\text{rot}}$ and $\mathcal{L}_{\text{flp}}$ are computed by comparing with the predicted angle θ in the original view (I_{ori}):

$$\mathcal{L}_{\text{rot}} = l_s(\theta_{\text{rot}} - \theta, R), \mathcal{L}_{\text{flp}} = l_s(\theta_{\text{flp}} + \theta, 0), \quad (14)$$

where l_s denotes a smooth L1 loss-based snap loss [48], and R denotes the angle applied to I_{ori}. The final angle loss is:

$$\mathcal{L}_{\text{ang}} = \beta(\lambda_r \mathcal{L}_{\text{rot}} + \lambda_f \mathcal{L}_{\text{flp}}) + \gamma \mathcal{L}_{\text{SPA}}, \quad (15)$$

where $\lambda_r = 1.0, \lambda_f = 0.05, \beta = 0.6$, and $\gamma = 0.05$ are empirically determined for all our experiments. In the shape learning branch (SLB), we use our IoU-based [45] regression loss $\mathcal{L}_{\text{reg}}$ in Eq. 11. The overall loss is defined as:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{ang}} \mathcal{L}_{\text{ang}} + \lambda_{\text{reg}} \mathcal{L}_{\text{reg}} + \lambda_{\text{cn}} \mathcal{L}_{\text{cn}} + \lambda_{\text{cls}} \mathcal{L}_{\text{cls}} \quad (16)$$

where $\mathcal{L}_{\text{cn}}$ is the center-ness loss [24] , and $\mathcal{L}_{\text{cls}}$ is the classification loss based on the focal loss [20]. The weighting factors, λ_{ang}, λ_{reg}, λ_{cn}, and λ_{cls} are all set to 1.

4. Experiments

4.1. Datasets

We trained and tested all the methods across four different datasets: DIOR [5, 13], DOTA-v1.0 [29], SIMD [9] and NWPU VHR-10 [4], which are summarized in Table 1. The details for the datasets and results for SIMD and NWPU are described in *Suppl*.

Datasets	# of Images	Image Widths	# of Objects	# of Classes	Annotation Types
DIOR [13]	22,463	800	190,288	20	T-HBox
DIOR-R [5]	22,463	800	190,288	20	RBox
DOTA-v1.0 [29]	2,806	800 $\sim$ 4K	188,282	15	C-HBox, RBox
SIMD [9]	5,000	1024	45,096	15	T-HBox
NWPU VHR-10 [4]	800	$\sim$1000	3,775	10	T-HBox

Table 1. Characteristics of datasets used for experiments

4.2. Implementation Details

Our proposed ABBSPO pipeline adopts the FCOS [24] detector as the baseline architecture, utilizing a ResNet-50 [10] backbone and an FPN [15] neck, based on the H2RBox-v2 [48] framework. To ensure fairness, all models are configured with the ResNet-50 [10] backbone and trained for 12 epochs on NVIDIA RTX3090 GPUs.

4.3. Experimental Results

4.3.1 Quantitative Comparison

It should be noted that objects such as round-shaped pools have *orientation ambiguities* regardless of their annotations (RBoxes) [42]. In order to avoid confusion in orientation learning, annotations are modified as having horizontal orientations if the objects belong to the following categories: (i) **DIOR-R**: 'baseball field', 'chimney', 'golf field', 'stadium', 'storage tank', 'windmill'; and (ii) **DOTA-v1.0**: 'baseball diamond', 'stadium', 'roundabout'. Accordingly, their orientation learning is enforced to predict the horizontal orientations, similar to previous works [42, 48].

Results on DIOR-R. Table 2 shows the OOD results. In addition to AP_{50} metric, we use 3-AP_{50} that focuses on the detection performance of the three complex-shaped object categories: 'airplane' (APL), 'expressway service area' (ESA), and 'overpass' (OP). As shown, our ABBSPO outperforms all weakly supervised OOD methods. Especially, in terms of 3-AP_{50}, our ABBSPO is superior to the HBox-supervised SOTA methods, H2RBox and H2RBox-v2, with large margins of average **12.9%**-point and average **9.1%**-point improvements. In overall AP_{50} performance, our ABBSPO surpasses H2RBox by **5.13%**-point and the H2RBox-v2 by **3.03%**-point. It is noted that our ABBSPO not only surpasses our base detector (H2RBox-v2 [48]) but also performs comparably to other RBox-supervised OOD methods, such as FCOS [24] and Oriented R-CNN [32]. It is worth noting that, compared to the RBox-supervised OOD methods, our ABBSPO shows *even* superior performance with *large* margins from **6.5**%-point to **11.7**%-point, especially on the 'airplane' that has the most complex shape. Notably, the ABBS module is less effective for rectangular objects, such as 'tennis court' (TC) and 'vehicle' (VE), as scaling is often unnecessary. However, it proves highly beneficial for complex-shaped objects, such as the ESA. The SPA loss is applied only to symmetric categories and helps

	Methods	APL	APO	BF	BC	BR	CH	ESA	ETS	DAM	GF	GTF	HA	OP	SH	STA	STO	TC	TS	VE	WM	3-AP$_{50}$	AP$_{50}$
$\mathcal{S}_R$	RetinaNet [20]	59.8	19.3	69.7	81.3	17.2	72.7	68.7	49.4	18.4	69.5	71.3	33.3	34.1	75.8	67.1	59.6	81.0	44.1	38.0	62.5	54.20	54.64
	FCOS [24]	62.1	37.9	74.6	81.2	32.9	72.1	75.3	61.8	27.4	69.1	78.7	34.4	50.6	80.1	68.6	68.1	81.3	49.1	43.4	64.5	62.67	60.66
	Oriented R-CNN [32]	63.0	36.7	71.9	81.6	41.1	72.6	77.8	65.5	24.8	72.9	82.1	40.9	56.5	81.2	73.4	62.4	81.5	53.3	43.3	65.6	65.77	62.41
	GWD [38] (RetinaNet)	61.5	23.6	73.6	81.1	17.4	72.7	68.3	47.2	20.7	71.2	73.2	33.9	34.3	77.6	64.7	57.5	80.9	42.1	39.7	60.2	54.70	55.07
	KLD [39] (RetinaNet)	57.8	22.6	71.5	81.2	16.9	72.7	68.9	52.1	20.6	73.5	71.0	33.7	33.2	77.1	68.9	59.9	80.9	43.9	39.1	60.9	53.30	55.32
	KFIoU [41] (RetinaNet)	60.6	36.6	73.6	80.9	27.0	72.6	73.4	56.5	25.4	73.9	72.0	32.9	45.8	75.8	65.2	57.6	80.0	48.0	40.1	58.8	59.93	57.84
$\mathcal{S}_I$	WSODet† [23]	20.7	29.0	63.2	67.3	0.2	65.5	0.4	0.1	0.3	49.0	28.9	0.3	1.5	1.2	53.4	16.4	40.0	0.1	6.1	0.1	7.53	22.20
$\mathcal{S}_P$	PointOBB† [17]	58.2	15.3	70.5	78.6	0.1	72.2	69.6	1.8	3.7	0.3	77.3	16.7	40.4	79.2	39.6	32.4	29.6	16.8	33.6	27.7	56.07	38.08
	Point2RBox-SK [47]	41.9	9.1	62.9	52.8	10.8	72.2	3.0	43.9	5.5	9.7	25.1	9.1	21.0	24.0	20.4	25.1	71.7	4.5	16.1	16.3	21.97	27.26
$\mathcal{S}_H$	H2RBox [42]	57.1	14.4	72.2	82.6	17.5	71.2	56.5	55.2	14	67.7	77.9	31	40.7	76.3	66.2	63.4	81.5	50.4	38	57.6	51.43	54.57
	H2RBox-v2 [48]	55.5	17.8	76.9	80.5	27.7	72.2	63.0	58.6	24.4	73.9	80.3	33.9	47.2	77.4	58.7	60.9	81.4	48.1	41.1	53.9	55.23	56.67
	ABBSPO (Ours)	69.5	15.7	76.2	87.5	29.9	72.3	75.3	61.2	28.1	74.1	81.7	34.7	48.2	79.3	67.4	61.4	81.5	54.7	41.5	53.8	**64.33**	**59.70**

Table 2. Quantitative results of each category on the DIOR-R [5] test dataset for RBox-supervised ($\mathcal{S}_R$), Image-supervised ($\mathcal{S}_I$), Point-supervised ($\mathcal{S}_P$) and HBox-supervised ($\mathcal{S}_H$) methods. The 3-AP$_{50}$ represents the mean AP$_{50}$ scores for three complex-shaped object categories: 'airplane' (APL), 'expressway service area' (ESA), and 'overpass' (OP). The notation † indicates its results in the paper [17].

	Methods	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	3-AP$_{50}$	AP$_{50}$
$\mathcal{S}_R$	RetinaNet [20]	87.5	75.1	39.9	59.6	66.3	66.3	78.2	90.5	55.0	62.7	47.1	63.6	59.4	55.1	43.0	61.87	63.3
	FCOS [24]	88.8	74.0	46.8	59.1	70.1	81.4	87.7	90.7	67.7	68.3	60.2	66.1	64.9	58.7	44.0	63.83	68.6
	Oriented R-CNN [32]	89.3	76.1	53.8	78.7	68.6	84.9	89.3	90.8	74.3	62.8	66.3	66.5	74.7	58.6	46.8	64.90	72.1
	Oriented RepPoints[14]	89.7	80.1	50.5	74.4	75.0	82.0	88.7	90.4	64.0	70.0	45.7	60.6	73.6	60.4	42.8	64.30	69.86
	GWD [38] (RetinaNet)	88.2	74.9	41.3	60.5	66.7	68.1	85.8	90.5	50.4	66.8	45.8	65.1	60.7	52.9	38.9	60.0	63.77
	KLD [39] (RetinaNet)	88.4	75.8	41.4	60.0	66.1	68.8	84.7	90.6	56.8	60.4	50.4	70.1	60.0	50.5	45.7	61.53	64.65
	KFIoU [41] (RetinaNet)	84.4	74.3	40.7	55.2	57.9	56.9	76.4	71.2	46.1	64.8	54.3	65.0	58.3	48.7	42.9	58.67	59.81
$\mathcal{S}_P$	PointOBB [17]+FCOS	32.4	67.3	0.8	53.6	2.3	9.7	18.8	0.3	9.9	12.8	0.5	54.0	11.0	34.1	11.4	25.97	21.26
	Point2RBox-SK [47]	50.1	63.7	1.6	44.7	23.9	34.7	32.7	78.8	41.2	32.2	2.1	34.3	20.8	42.5	7.2	33.27	34.03
$\mathcal{S}_H$	H2RBox [42]	89.5	73.1	37.3	55.1	70.7	76.4	85.4	90.3	66.5	67.3	59.6	64.9	60.6	57.9	36.5	61.30	66.07
	H2RBox-v2 [48]	89.4	74.8	45.4	56.0	70.3	76.6	87.9	90.5	69.3	67.5	56.7	64.7	65.3	55.5	45.5	63.47	67.69
	ABBSPO (Ours)	89.2	75.6	47.4	52.8	70.3	77.6	88.2	90.5	67.9	66.8	68.2	66.2	71.6	55.6	51.0	**65.27**	**69.26**

Table 3. Quantitative results of each category on the DOTA-v1.0 [29] validation dataset for $\mathcal{S}_R$, $\mathcal{S}_I$, $\mathcal{S}_P$ and $\mathcal{S}_H$ methods. The 3-AP$_{50}$ represents the mean AP$_{50}$ scores for three complex-shaped object categories: plane (PL), swimming pool (SP), and helicopter (HC). All the methods are re-trained using only train dataset for fair comparison.

improve their performance, except for the categories with *orientation ambiguities*, such as 'storage tank' (STO). Since the predicted angles learned through the SPA loss are also utilized in the ABBS module for scale adjustment, both the SPA loss and the ABBS module jointly contribute to performance improvement in symmetric categories. This joint effect is particularly evident in complex-shaped symmetric categories, such as APL and ESA, where performance gains are more significant. Nevertheless, the performance gains for the two symmetric and rectangular categories, TC and VE, are marginal. This is mainly because the ABBS module has limited impact on rectangular shapes, and the small object sizes lead to an insufficient number of pixels for reliably determining the symmetry axis via the SPA loss.

Results on DOTA-v1.0. Table 3 shows the detection performance results on the DOTA-v1.0 [29]. Due to the non-responsiveness of the DOTA evaluation server, we report our experimental results on the validation dataset (458 images) instead of the test dataset (937 images). It should be noted that the validation dataset was not used for training all the methods for fair comparison. We use 3-AP$_{50}$ that measures the detection performance for the three complex-shaped object categories: 'plane', 'swimming pool' and 'helicopter'. Our ABBSPO achieves SOTA performance, outperforming H2RBox by **3.19%**-point and H2RBox-v2 by **1.57%**-point improvements. Moreover, our ABBSPO even surpasses the FCOS Baseline by **0.66%**-point lift.

4.3.2 Qualitative Comparison

Results on DIOR. As shown in the figures of the first row in Fig. 5, our ABBSPO is the only method that accurately captures both the orientation and scale of the airplane. Since DIOR annotations provide GT in T-HBox format, direct usage of T-HBoxes as GT for training to predict RBox leads to degradations in orientation and scale prediction accuracy for the existing HBox-supervised OOD methods as shown in the figures of columns 2 and 3 in Fig. 5. In contrast, our ABBSPO avoids such degradation by utilizing the ABBS module that optimally scales the GT HBox sizes for precise RBox prediction during training. It is also worthwhile to mention that the predicted orientations by our ABBSPO are more precisely obtained via our SPA loss. Furthermore, it should be noted that compared to the RBox-supervised baseline method (Rotated FCOS [24]), our approach demonstrates superior visual results, even under weakly supervised learning.

Results on DOTA-v1.0. As shown in the figures of the second row in Fig. 5, ABBSPO very accurately predicts both the orientation and scale of the swimming pool, achieving similar accuracy for tennis court. Interestingly, only ABBSPO successfully detects the two tennis courts that are partially occluded by trees (red solid circle) while the other

Module		DIOR-R		DOTA-v1.0
ABBS	SPA	3-AP_{50}	AP_{50}	AP_{50}
		55.23	56.67	67.69
✓		62.13	58.35	68.59
	✓	58.77	58.99	69.16
✓	✓	**64.33**	**59.70**	**69.26**

Table 4. Ablation results on ABBS module and SPA loss ($\mathcal{L}_{\text{SPA}}$).

Sampling		DIOR-R	
$\mathcal{L}_{\text{SPA}}$	Others	3-AP_{50}	AP_{50}
		61.67	58.93
✓		**64.33**	**59.70**
	✓	43.63	50.51
✓	✓	45.1	50.91

Table 5. Ablation results on proposal sampling in $\mathcal{L}_{\text{SPA}}$ and other components.

Scale Range			DIOR-R		DOTA-v1.0
Min	Max	Interval	3-AP_{50}	AP_{50}	AP_{50}
0.9	1.1	0.05	57.97	58.15	**69.26**
0.5	1.5	0.1	61.67	59.62	68.8
1.0	1.5	0.1	**64.33**	**59.70**	68.9
1.0	2.0	0.1	56.07	55.46	66.55

Table 6. Ablation results on scale range in ABBS module.

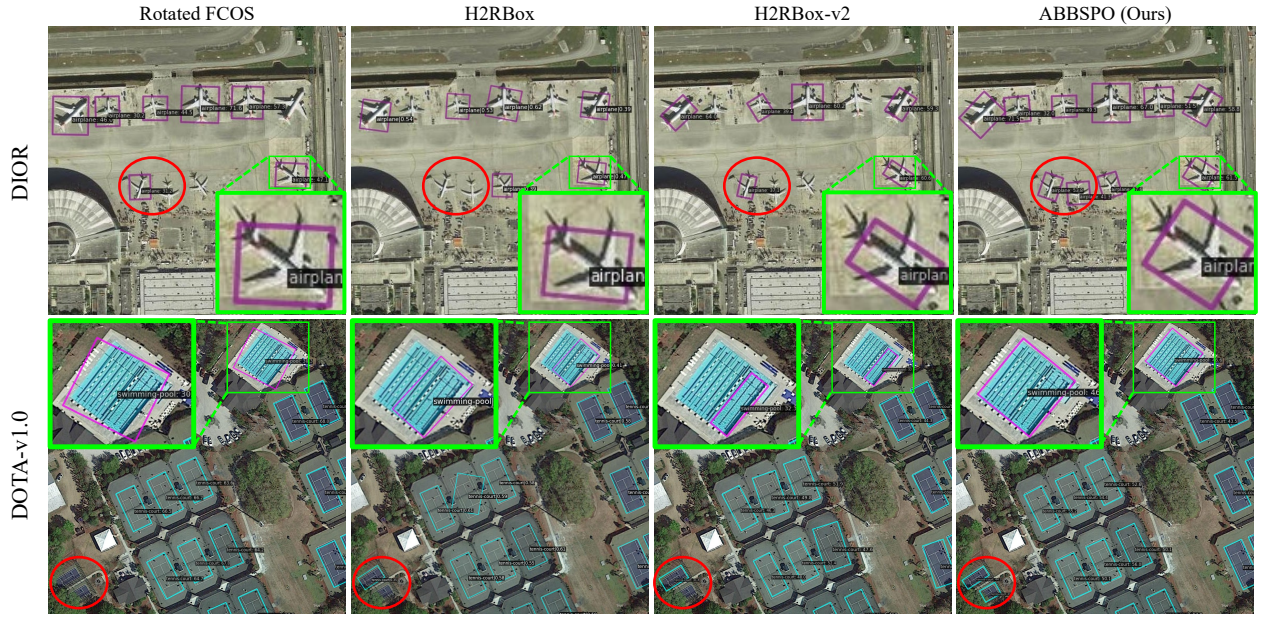


Figure 5. Qualitative results on DIOR [5, 13] and DOTA-v1.0 [29] . Zoom-in for better visualization. Rotated FCOS was trained only with GT RBoxes, while H2RBox, H2RBox-v2 and our ABBSPO were trained with GT T-HBoxes (1st row) and GT C-HBoxes (2nd row).

methods failed. These results visually support the effectiveness of our ABBS module and SPA loss in learning the scales and orientations of objects accurately.

4.4. Ablation Studies

Ablation study on SPA loss and ABBS module. As shown in Table 4, both components contribute to performance improvements. The ABBS module effectively scales the GT HBoxes, leading to an increase in AP_{50} performance on the DIOR dataset. Notably, it has a greater effect on complex-shaped object categories, resulting in a significant improvement in 3-AP_{50}. Similarly, the SPA loss enhances angle prediction accuracy, also bringing an improvement in AP_{50}.

Ablation study on proposal sampling. As shown in Table 5, applying Top-k proposal sampling exclusively to the SPA loss ($\mathcal{L}_{\text{SPA}}$) yields the highest AP_{50} performance, as the symmetric proposals of high-quality benefits $\mathcal{L}_{\text{SPA}}$. But, additional proposal sampling to the others ($\mathcal{L}_{\text{rot}}$, $\mathcal{L}_{\text{flp}}$, $\mathcal{L}_{\text{reg}}$, $\mathcal{L}_{\text{cn}}$, $\mathcal{L}_{\text{cls}}$) significantly lowers the performance.

Ablation study on scale range in the ABBS module. As shown in Table 6, the optimal scale range is influenced by the type of GT HBoxes. For DIOR's T-HBoxes, a scale range of 1 to 1.5 works well because it ensures that the predicted RBoxes fully cover the objects boundary. On the other hand, for DOTA's C-HBoxes, which are already close to the optimal HBoxes, the optimal scale range is closer to 1. By adjusting the scale range based on the type of HBoxes, the ABBS module achieves high accuracy in predicting RBoxes for both datasets.

5. Conclusion

Our ABBSPO, a weakly supervised OOD framework, effectively learns RBox prediction regardless of the type of HBox annotations (T-HBox and C-HBox). With our proposed Adaptive Bounding Box Scaling (ABBS) and Symmetric Prior Angle (SPA) loss, we achieved enhanced orientation and scale accuracy for OOD, which is comparable to or even better than RBox-supervised methods. Extensive experimental results underscore the superiority of our approach, surpassing state-of-the-art HBox-supervised methods. Our method effectively bridges the gap between weakly supervised OOD and fully supervised OOD, making it a promising solution for applications requiring efficient and accurate object detection via training with relatively cheap annotations of HBoxes compared to RBoxes.

Acknowledgement

This research was supported by Korea Institute of Marine Science & Technology Promotion (KIMST) funded by the Korea Coast Guard (RS-2023-00238652, Integrated Satellite-based Applications Development for Korea Coast Guard, 100%).

References

[1] Liangyu Chen, Tong Yang, Xiangyu Zhang, Wei Zhang, and Jian Sun. Points as queries: Weakly semi-supervised object detection by points. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8823–8832, 2021. 3

[2] Pengfei Chen, Xuehui Yu, Xumeng Han, Najmul Hassan, Kai Wang, Jiachen Li, Jian Zhao, Humphrey Shi, Zhenjun Han, and Qixiang Ye. Point-to-box network for accurate object detection via single point supervision. In *European Conference on Computer Vision*, pages 51–67. Springer, 2022. 3

[3] Yihong Chen, Zheng Zhang, Yue Cao, Liwei Wang, Stephen Lin, and Han Hu. Reppoints v2: Verification meets regression for object detection. In *Advances in Neural Information Processing Systems*, pages 5621–5631. Curran Associates, Inc., 2020. 2

[4] Gong Cheng, Peicheng Zhou, and Junwei Han. Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images. *IEEE transactions on geoscience and remote sensing*, 54(12):7405–7415, 2016. 2, 6

[5] 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. 6, 7, 8

[6] Jian Ding, Nan Xue, Yang Long, Gui-Song Xia, and Qikai Lu. Learning roi transformer for oriented object detection in aerial images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2849–2858, 2019. 2

[7] 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, 2021. 2

[8] 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*, pages 2786–2795, 2021. 2

[9] Muhammad Haroon, Muhammad Shahzad, and Muhammad Moazam Fraz. Multisized object detection using spaceborne optical imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:3032–3046, 2020. 2, 6

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6

[11] Shitian He, Huanxin Zou, Yingqian Wang, Boyang Li, Xu Cao, and Ning Jing. Learning remote sensing object detection with single point supervision. *IEEE Transactions on Geoscience and Remote Sensing*, 2023. 3

[12] Darius Lam, Richard Kuzma, Kevin McGee, Samuel Dooley, Michael Laielli, Matthew Klaric, Yaroslav Bulatov, and Brendan McCord. xview: Objects in context in overhead imagery. *arXiv preprint arXiv:1802.07856*, 2018. 2

[13] Ke Li, Gang Wan, Gong Cheng, Liqiu Meng, and Junwei Han. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS journal of photogrammetry and remote sensing*, 159:296–307, 2020. 2, 6, 8

[14] Wentong Li, Yijie Chen, Kaixuan Hu, and Jianke Zhu. Oriented reppoints for aerial object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1829–1838, 2022. 2, 7

[15] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017. 6

[16] Xuebo Liu, Ding Liang, Shi Yan, Dagui Chen, Yu Qiao, and Junjie Yan. Fots: Fast oriented text spotting with a unified network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5676–5685, 2018. 3

[17] Junwei Luo, Xue Yang, Yi Yu, Qingyun Li, Junchi Yan, and Yansheng Li. Pointobb: Learning oriented object detection via single point supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16730–16740, 2024. 2, 3, 7

[18] Botao Ren, Xue Yang, Yi Yu, Junwei Luo, and Zhidong Deng. Pointobb-v2: Towards simpler, faster, and stronger single point supervised oriented object detection. *arXiv preprint arXiv:2410.08210*, 2024. 3

[19] Zhongzheng Ren, Zhiding Yu, Xiaodong Yang, Ming-Yu Liu, Alexander G Schwing, and Jan Kautz. Ufo 2: A unified framework towards omni-supervised object detection. In *European conference on computer vision*, pages 288–313. Springer, 2020. 3

[20] T-YLPG Ross and GKHP Dollár. Focal loss for dense object detection. In *proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2980–2988, 2017. 2, 6, 7

[21] Xian Sun, Peijin Wang, Zhiyuan Yan, Feng Xu, Ruiping Wang, Wenhui Diao, Jin Chen, Jihao Li, Yingchao Feng, Tao Xu, et al. Fair1m: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184:116–130, 2022. 2

[22] Yongqing Sun, Jie Ran, Feng Yang, Chenqiang Gao, Takayuki Kurozumi, Hideaki Kimata, and Ziqi Ye. Oriented object detection for remote sensing images based on weakly supervised learning. In *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–6. IEEE, 2021. 2

[23] Zhiwen Tan, Zhiguo Jiang, Chen Guo, and Haopeng Zhang. Wsodet: A weakly supervised oriented detector for aerial object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–12, 2023. 2, 3, 7

[24] Zhi Tian, Xiangxiang Chu, Xiaoming Wang, Xiaolin Wei, and Chunhua Shen. Fully convolutional one-stage 3d object detection on lidar range images. *Advances in Neural Information Processing Systems*, 35:34899–34911, 2022. 2, 4, 6, 7

[25] Hao Wang, Zhanchao Huang, Zhengchao Chen, Ying Song, and Wei Li. Multigrained angle representation for remote-sensing object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022. 2

[26] Jian Wang, Fan Li, and Haixia Bi. Gaussian focal loss: Learning distribution polarized angle prediction for rotated object detection in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022. 2

[27] Linfei Wang, Yibing Zhan, Xu Lin, Baosheng Yu, Liang Ding, Jianqing Zhu, and Dapeng Tao. Explicit and implicit box equivariance learning for weakly-supervised rotated object detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024. 2

[28] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6

[29] 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*, pages 3974–3983, 2018. 2, 6, 7, 8

[30] SUN Xian, WANG Zhirui, SUN Yuanrui, DIAO Wenhui, ZHANG Yue, and FU Kun. Air-sarship-1.0: High-resolution sar ship detection dataset. , 8(6):852–863, 2019. 2

[31] Zhifeng Xiao, Qing Liu, Gefu Tang, and Xiaofang Zhai. Elliptic fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images. *International Journal of Remote Sensing*, 36(2):618–644, 2015. 2

[32] 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*, pages 3520–3529, 2021. 6, 7

[33] 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*, pages 17406–17415, 2024. 2

[34] Xue Yang and Junchi Yan. Arbitrary-oriented object detection with circular smooth label. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, pages 677–694. Springer, 2020. 2

[35] Xue Yang and Junchi Yan. Arbitrary-oriented object detection with circular smooth label. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, pages 677–694. Springer, 2020. 2

[36] Xue Yang, Liping Hou, Yue Zhou, Wentao Wang, and Junchi Yan. Dense label encoding for boundary discontinuity free rotation detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15819–15829, 2021. 2

[37] Xue Yang, Junchi Yan, Ziming Feng, and Tao He. R3det: Refined single-stage detector with feature refinement for rotating object. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(4):3163–3171, 2021. 2

[38] Xue Yang, Junchi Yan, Qi Ming, Wentao Wang, Xiaopeng Zhang, and Qi Tian. Rethinking rotated object detection with gaussian wasserstein distance loss. In *International conference on machine learning*, pages 11830–11841. PMLR, 2021. 2, 7

[39] Xue Yang, Xiaojiang Yang, Jirui Yang, Qi Ming, Wentao Wang, Qi Tian, and Junchi Yan. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence. *Advances in Neural Information Processing Systems*, 34:18381–18394, 2021. 7

[40] Xue Yang, Gefan Zhang, Xiaojiang Yang, Yue Zhou, Wentao Wang, Jin Tang, Tao He, and Junchi Yan. Detecting rotated objects as gaussian distributions and its 3-d generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4335–4354, 2022.

[41] 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*, 2022. 2, 7

[42] Xue Yang, Gefan Zhang, Wentong Li, Yue Zhou, Xuehui Wang, and Junchi Yan. H2RBox: Horizontal box annotation is all you need for oriented object detection. In *The Eleventh International Conference on Learning Representations*, 2023. 1, 2, 3, 5, 6, 7

[43] Ze Yang, Shaohui Liu, Han Hu, Liwei Wang, and Stephen Lin. Reppoints: Point set representation for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9657–9666, 2019. 2

[44] Xinyi Ying, Li Liu, Yingqian Wang, Ruojing Li, Nuo Chen, Zaiping Lin, Weidong Sheng, and Shilin Zhou. Mapping degeneration meets label evolution: Learning infrared small target detection with single point supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15528–15538, 2023. 3

[45] Jiahui Yu, Yuning Jiang, Zhangyang Wang, Zhimin Cao, and Thomas Huang. Unitbox: An advanced object detection network. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 516–520, 2016. 6

[46] 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*, pages 13354–13363, 2023. 4

[47] Yi Yu, Xue Yang, Qingyun Li, Feipeng Da, Jifeng Dai, Yu Qiao, and Junchi Yan. Point2rbox: Combine knowledge from synthetic visual patterns for end-to-end oriented object detection with single point supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16783–16793, 2024. 2, 3, 7

[48] Yi Yu, Xue Yang, Qingyun Li, Yue Zhou, Feipeng Da, and Junchi Yan. H2rbox-v2: Incorporating symmetry for boost-

ing horizontal box supervised oriented object detection. *Advances in Neural Information Processing Systems*, 36, 2024. 1, 2, 3, 4, 5, 6, 7

[49] Tingxuan Yue, Yanmei Zhang, Jin Wang, Yanbing Xu, and Pengyun Liu. A weak supervision learning paradigm for oriented ship detection in sar image. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 2