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

Supplementary Material

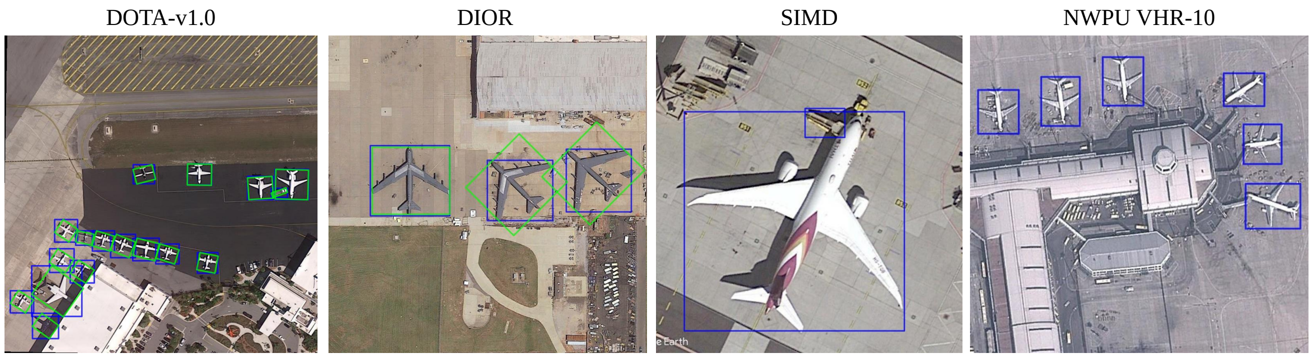


Figure 6. Visualization of GT Box Types for Different Datasets.

In this supplementary material, we present a detailed description and visualization of the datasets we used for training and evaluation, including DOTA-v1.0 [29], DIOR [13], DIOR-R [5], SIMD [9], and NWPU VHR-10 [4] datasets. Furthermore, we validate the effectiveness of our methods through two additional ablation studies on (i) the scale adjustment function and (ii) the scale ranges for the DOTA-v1.0 dataset. Also, we provide quantitative and qualitative results on DIOR-R [5] under the experimental settings based on the H2RBox [42]'s public source codes. Lastly, we provide additional qualitative results on DIOR [5, 13], DOTA-v1.0 [29], SIMD [9], and NWPU VHR-10 [4] datasets.

A. Details of Datasets

A.1. Detailed Description of Datasets

DOTA-v1.0 [29] comprises 2,806 images with 188,282 annotated instances across 15 categories, having both rotated bounding boxes (GT RBoxes) and coarse horizontal bounding boxes (GT C-HBoxes) for annotation. Among these images, 1,411 are designated for training, 458 for validation, and 937 for testing. The image dimensions range from 800×800 to $4,000 \times 4,000$. During training, the images were cropped into a patch size of $1,024 \times 1,024$. The dataset includes 15 categories: 'plane' (PL), 'baseball-diamond' (BD), 'bridge' (BR), 'ground-track-field' (GTF), 'small-vehicle' (SV), 'large-vehicle' (LV), 'ship' (SH), 'tennis-court' (TC), 'basketball-court' (BC), 'storage-tank' (ST), 'soccer-ball-field' (SBF), 'roundabout' (RA), 'harbor' (HA), 'swimming-pool' (SP), and 'helicopter' (HC).

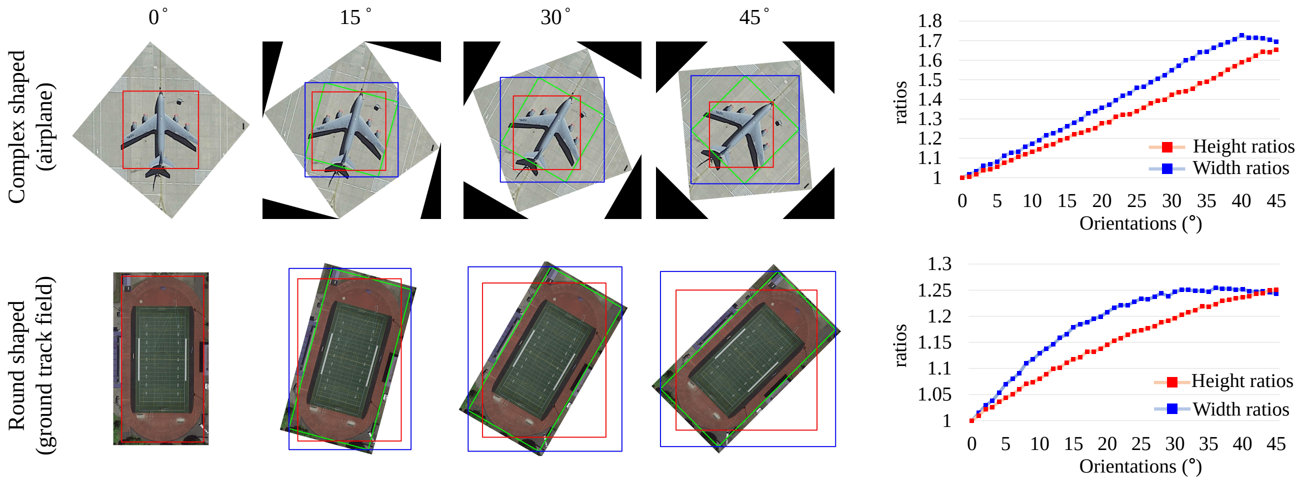
In our experimental setup, all methods including our ABSSPO are trained on the train split of DOTA-v1.0 [29] with GT C-HBoxes and evaluated on its validation split with GT RBoxes.

DIOR [13] contains 800×800-sized 23,463 aerial images of 20 categories with 190,288 instances (objects), each having tight horizontal bounding box annotations (GT T-HBoxes). Among these, 5,862 images are used for training, 5,863 images for validation, and the remaining 11,738 images for testing. The 20 categories of the dataset include: 'airplane' (APL), 'airport' (APO), 'baseball field' (BF), 'basketball court' (BC), 'bridge' (BR), 'chimney' (CH), 'expressway service area' (ESA), 'expressway toll station' (ETS), 'dam' (DAM), 'golf field' (GF), 'ground track field' (GTF), 'harbor' (HA), 'overpass' (OP), 'ship' (SH), 'stadium' (STA), 'storage tank' (STO), 'tennis court' (TC), 'train station' (TS), 'vehicle' (VE), and 'windmill' (WM).

DIOR-R [5] contains the same images as DIOR [13], but includes rotated bounding box annotations for its objects instead of HBox annotations. It should be noted that in our experiments, we utilize both the train and validation splits in DIOR [13] dataset for training our method, while employing the test split in DIOR-R [5] dataset for evaluation.

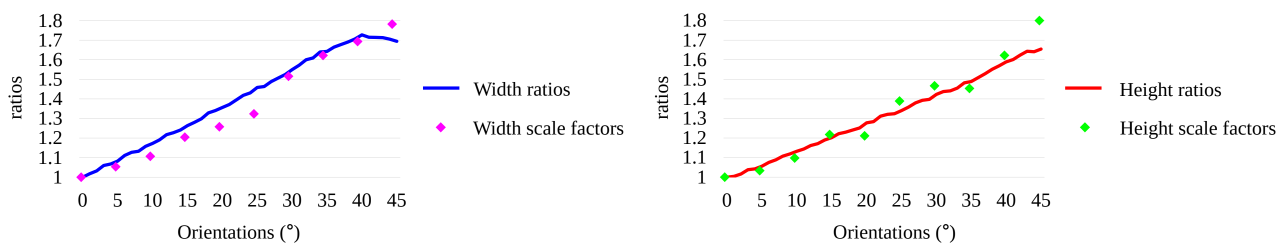
SIMD [9] comprises aerial images annotated with tight horizontal bounding boxes (GT T-HBoxes). The dataset contains 5,000 images spanning 15 categories, with a total of 45,096 instances. Each image has a fixed width of 1,024 pixels and fixed-sized heights of 768 pixels. The 15 categories are as follows: 'car', 'truck', 'van', 'longvehicle', 'bus', 'airliner', 'propeller', 'trainer', 'chartered', 'fighter', 'other', 'stairtruck', 'pushbacktruck', 'helicopter', and 'boat'.

NWPU VHR-10 [4] is an aerial image dataset featuring tight horizontal bounding box (GT T-HBox) annotations. It



(a) Variations of T-HBoxes, RBoxes, and derived HBoxes for orientations (rotation angles)   (b) Ratios b/w T-HBoxes and derived HBoxes

Figure 7. Analysis on variations of width and height ratios between the T-HBoxes and the HBoxes derived as minimum circumscribed rectangles of RBoxes. (a) Variations in the shapes of manually annotated T-HBoxes, manually annotated RBoxes and derived HBoxes from the RBoxes for various rotations angles of different objects (airplanes in the first row and ground track fields in the second row); (b) Width and height ratios between T-HBoxes and corresponding derived HBoxes for the rotation angles of the airplanes (top) and the ground track fields (bottom).



(a) Comparison between width ratios and width scale factors   (b) Comparison between height ratios and height scale factors

Figure 8. Comparisons between width (height) ratios and scale factors. (a) Width ratio curve (blue curve) calculated from manual annotations, and *angle-adjusted* width scale factors (diamond-shaped pink points) derived from best scaled HBoxes selected by the ABBS; (b) height ratio curve (red curve) calculated from manual annotations and *angle-adjusted* height scale factors (diamond-shaped green points) derived from best scaled HBoxes selected by the ABBS. It is noted that the *angle-adjusted* width (height) scale factors are well aligned with the width (height) curves, indicating that our ABBS works properly in scaling the GT HBoxes.

comprises 800 images spanning 10 categories, with approximately 3,775 annotated instances. The images have widths of around 1,000 pixels. The dataset includes the following 10 categories: 'airplane', 'ship', 'storage tank', 'baseball diamond', 'tennis court', 'basketball court', 'ground track field', 'harbor', 'bridge', and 'vehicle'.

After training on the SIMD [9] and NWPU VHR-10 [4] datasets, we focus on qualitative comparisons only, as these datasets do not provide RBox annotations.

A.2. Visualization of GT Box Types for Different Datasets.

Fig. 6 illustrates how GT HBoxes are annotated across different datasets. In this figure, the GT RBoxes are marked in green, while the GT HBoxes are marked in blue. For the SIMD [9] and NWPU VHR-10 datasets [4] where their GT RBoxes do not exist, only the GT HBoxes (blue) are displayed. The comparison emphasizes airplanes that have complex shapes, to highlight the differences in the GT HBox annotation types.

DOTA-v1.0 [29]. The GT HBoxes (blue) in DOTA are annotated as the minimum circumscribed HBoxes for their corresponding GT RBoxes (green), as can be seen in the first column of Fig. 6. For the objects with larger rotation angles, the sizes of their GT HBoxes (blue) appear larger-sized, when being more apart from the objects' boundaries. Such GT HBoxes (blue) that are derived directly from the objects boundaries were previously defined as GT C-

HBoxes in Sec. 1 of the main paper.

DIOR [13] (& **DIOR-R** [5]). Its GT HBoxes (blue) are sourced from the DIOR dataset, while the GT RBoxes (green) are taken from the DIOR-R dataset for visualization purposes. The GT HBoxes (blue) tightly enclose the objects' boundaries, and are annotated independently from the GT RBoxes (green). This type of HBox annotations is referred to as GT T-HBoxes in the main paper. As the orientations of the objects increase, their GT RBoxes (green) extend further beyond their corresponding GT HBoxes (blue), which can be observed in the center and right airplanes in the second column of Fig. 6.

DOTA [29] vs. **DIOR** [13] (& **DIOR-R** [5]). In the DOTA and DIOR datasets, their objects are annotated as GT C-HBoxes and GT T-HBoxes, respectively. The GT HBoxes (blue) and GT RBoxes (green) align perfectly each other when the objects' orientations are horizontal or vertical, as demonstrated by the top-right airplanes in the first column of Fig. 6 and the left airplane in the second column of Fig. 6. However, as the objects get more rotated from the horizontal or vertical angle, their GT C-HBoxes (blue) become consistently enlarged to circumscribe their corresponding GT RBoxes (green), as seen in the bottom-left airplanes in the first column of Fig. 6. In contrast, as depicted in the second column of Fig. 6, the GT T-HBoxes always tightly enclose the objects boundaries, regardless of their GT RBoxes (green). This difference in GT HBox annotation leads to a significant degradation in the OOD performance of the previous HBox-supervised OOD methods when trained on the DIOR dataset with GT T-HBoxes, unlike when trained on the DOTA dataset with GT C-HBoxes.

SIMD [9] & **NWPU VHR-10** [4]. As shown in the third and forth columns of Fig. 6, SIMD and NWPU VHR-10 datasets contains only GT T-Hboxes, where the GT T-HBoxes (blue) tightly enclose the boundaries of airplanes, even for objects with large orientation angles.

B. Additional Ablation Study

Ablation Study on Scale Adjustment Function. As mentioned in Sec. 3.2 of the main paper, we incorporate the object shape types and orientation degrees into the scale adjustment of the widths and heights of GT T-HBoxes. The scale adjustment function f is designed as a linear function of the angle θ, as presented in Eq. 8 of the main paper.

Fig. 7-(a) demonstrates this process using two images containing objects from the 'airplane' class and 'ground track field' class, which are manually rotated from 0° to 45° in 15° increments. On the top of each rotated image, a T-HBox (red), an RBox (green), and a minimum circumscribed HBox (blue) derived from the RBox are overlayed. Fig. 7-(b) shows the width and height ratios between the T-HBoxes and their derived HBoxes. As shown, the results indicate that these ratios increase linearly as the orientation degrees increase, validating that the proposed linear scale adjustment function f can effectively capture the scale variations in annotations caused by the object orientations.

Additionally, Fig. 8 visualizes the scale proportion between the annotated T-HBoxes and their best scaled HBoxes in the ABBS module, by passing manually rotated airplane images from 0° to 45°. This scale proportion corresponds to the *angle-adjusted* scale factors derived from Eq. 7, Eq. 8, and Eq. 10 in the main paper, which are represented as diamond-shaped points in Fig. 8. Fig. 8-(a) compares the width ratio curve (blue curve) calculated from manually annotated T-HBoxes and derived HBoxes from manually annotated RBoxes with the *angle-adjusted* width scale factors (diamond-shaped pink points) obtained from best scaled HBoxes selected by the ABBS. Fig. 8-(b) compares the height ratio curve (red curve) calculated from manually annotated T-HBoxes and derived HBoxes from manually annotated RBoxes with the *angle-adjusted* height scale factors (diamond-shaped green points) obtained from best scaled HBoxes selected by the ABBS. It is noted in Fig. 8-(a) and -(b) that the *angle-adjusted* width (height) scale factors are well aligned with the width (height) curves, indicating that our ABBS works properly in scaling the GT HBoxes. This alignment highlights the effectiveness of our proposed ABBS module in capturing and leveraging the scale variations caused by the object orientation, enabling precise adaptation to changes in object orientations and shapes during training.

Scale Range			DOTA-v1.0	
Min	Max	Interval	3-AP_{50}	AP_{50}
0.9	1.0	0.05	61.80	68.09
1.0	1.1	0.05	64.77	69.08
0.9	1.1	0.05	**65.27**	**69.26**
0.8	1.1	0.05	63.33	69.03

Table 7. Ablation results on the scale ranges of GT HBoxes for ABBS module in the DOTA-v1.0 [29] dataset.

Ablation Study on Scale Ranges of GT HBoxes for ABBS Module. Our ABBS module adjusts the sizes of given GT HBoxes during the training, with their scale adjustment range depending on the annotation types of the datasets. Especially for the DIOR dataset [13] that uses GT T-HBoxes, the scale range is set from 1 to 1.5 to optimize the training process as shown in Table 6 of the main paper. Conversely, for the DOTA-v1.0 dataset [29] that provides GT C-HBoxes, a narrower scale range of 0.9 to 1.1 is employed. As shown in Table 7, the best OOD performance on DOTA-v1.0 dataset is achieved when the scale range is set between 0.9 and 1.1, validating the necessity of dataset-specific scale adjustments.

Methods	<u>APL</u>	APO	BF	BC	BR	CH	<u>ESA</u>	ETS	DAM	GF	GTF	HA	<u>OP</u>	SH	STA	STO	TC	TS	VE	WM	<u>3-AP$_{50}$</u>	AP$_{50}$
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* [42]	65.5	12.5	74.6	81.3	21.3	72.2	62.7	60.4	19.2	70.1	78.7	35.3	44.3	79.1	62.7	68.2	81.5	51.7	39.6	60.7	57.50	57.08
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
H2RBox-v2* [48]	67.2	11.5	75.8	84.0	31.4	72.5	65.3	60.7	25.3	72.2	80.9	35.2	50.2	78.9	67.0	61.5	81.5	52.6	43.0	26.7	60.90	57.17
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**
ABBSPO (Ours)*	66.6	20.2	77.6	84.7	30.8	72.5	75.0	60.1	28.3	75.3	81.2	35.9	49.0	79.4	69.7	65.3	81.4	55.1	41.7	33.0	**63.53**	**59.14**

Table 8. Quantitative OOD results for various object categories on the DIOR-R [5] dataset. Results marked with * are obtained by running H2RBox [42]'s public source codes. For these results, the experiment settings do not include angle prediction for the airplane class. Results without * are obtained using the original configuration that is the same as that in the main paper.

Figure 9. Qualitative OOD comparisons on DIOR [5, 13]. Results marked with * are obtained by running H2RBox [42]'s public source codes. For these results, the experiment settings do not include angle prediction for the airplane category. Results without * are obtained using the original configuration that is the same as that in the main paper.

C. Additional Results on DIOR

As discussed in Sec. 4.3.1 of the main paper, our ABBSPO demonstrates superior OOD performance compared to the previous methods, H2RBox [42] and H2RBox-v2 [48]. For fair comparison, we set the objects belonging to the following six classes ('baseball field', 'chimney', 'golf field', 'stadium', 'storage tank', and 'windmill') as the subjects not to predict their orientations due to orientation ambiguities. We denote this setting as the original configuration that is the same as that in the main paper. We further provide the OOD results under an additional configuration, where six different classes ('airplane', 'baseball field', 'chimney', 'golf field', 'stadium', and 'storage tank') were designated as objects without orientation prediction, following the H2Rox [42]'s public source codes.

We denote the methods trained and evaluated in the above additional configuration as H2RBox* [42], H2RBox-v2* [48], and ABBSPO*, while H2RBox [42], H2RBox-v2 [48], and ABBSPO denote the methods trained and evaluated in the original configuration. As illustrated in Fig. 9, H2RBox* [42], H2RBox-v2* [48], and ABBSPO* predict objects in 'airplane' class in the form of HBoxes without orientations, while H2RBox [42], H2RBox-v2 [48], and ABBSPO predicts orientations of the objects for the same class. As shown in Table 8, our ABBSPO* still outperforms H2RBox* [42] and H2RBox-v2* [48]. These findings demonstrate the effectiveness and robustness of our ABBSPO over different configurations for the orientation ambiguity.

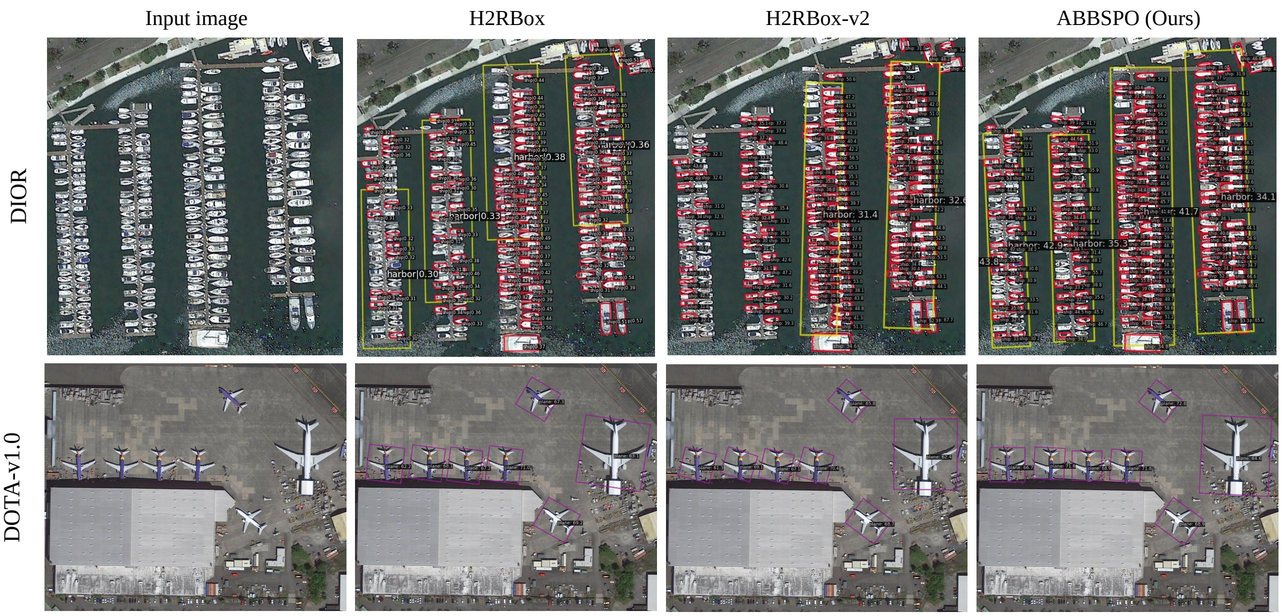


Figure 10. Qualitative OOD results on DIOR [5, 13] and DOTA-v1.0 [29] datasets.

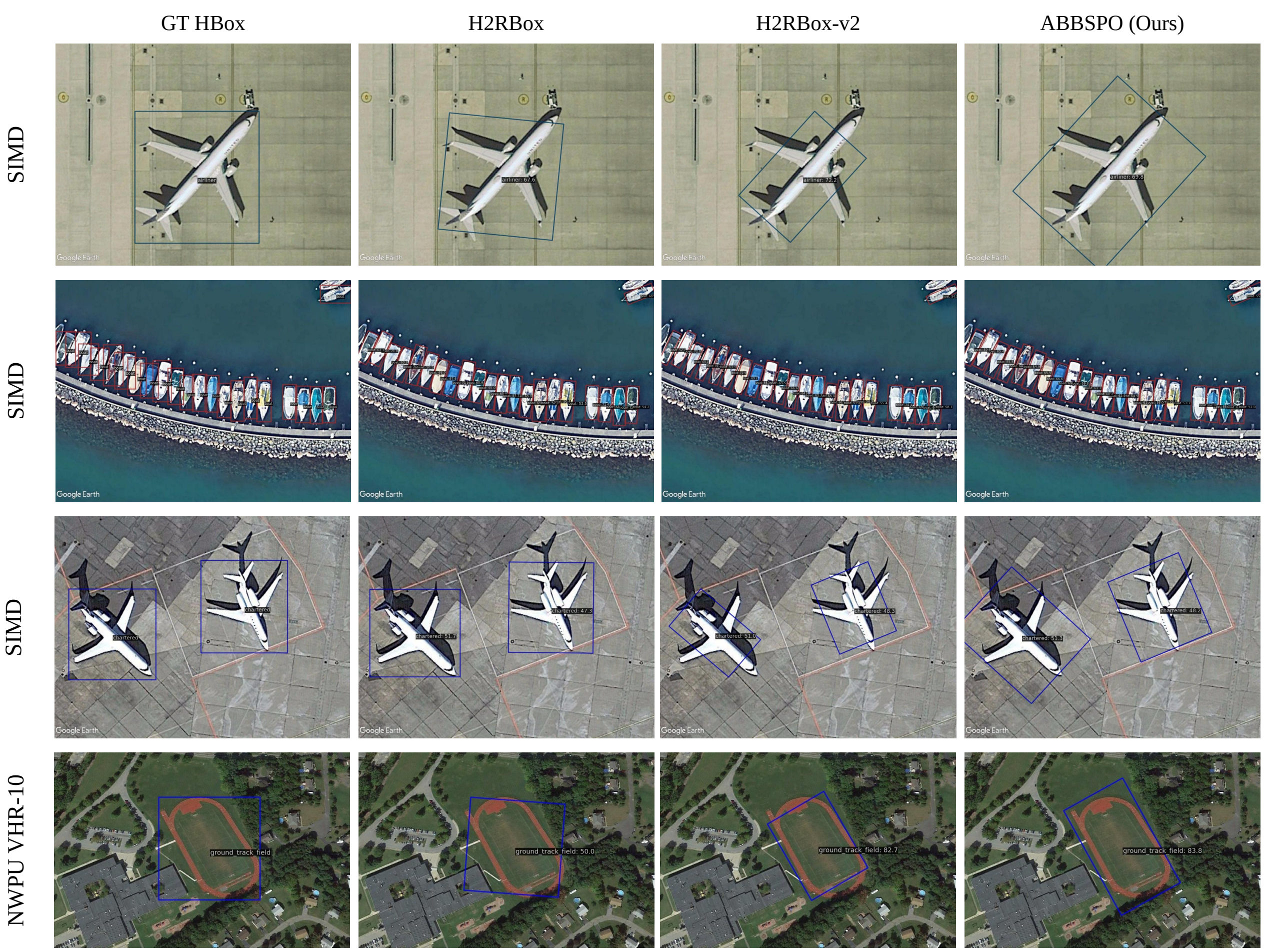


Figure 11. Qualitative OOD results on SIMD [9] and NWPU VHR-10 [4] datasets.

D. Additional Qualitative Results

DIOR. As shown in the first row of Fig. 10, our ABBSPO successfully detects more ships and harbors with higher precision. This demonstrates an improved object capturing ability of our model, achieved through an effective training process.

DOTA-v1.0. As shown in the second row of Fig. 10, our ABBSPO predicts the orientations of airplanes more accurately. This highlights an enhanced angle prediction accuracy of our ABBSPO, which is reinforced by our SPA loss-based self-supervision during training.

SIMD & NWPU VHR-10. As shown in the first column of Fig. 11, the GT HBoxes for the corresponding test images are displayed since they only contain GT T-HBoxes. Across the four test images, our ABBSPO demonstrates superior OOD performance in terms of both scale and orientation predictions. Notably, in the first and third rows that show predictions for airplanes, our ABBSPO successfully predicts accurate orientations and scales of the objects while H2RBox [42] and H2RBox-v2 [48] predicts inaccurate orientations and scales for the objects, which often occurs when their models are trained with GT T-HBoxes.

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, 1, 3, 5

[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, 1, 3, 4, 5

[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, 1, 3, 5

[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, 1, 3, 4, 5

[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, 1, 3, 5

[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, 4

[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 boosting 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