

# DARDet: A Dense Anchor-free Rotated Object Detector in Aerial Images

Feng Zhang, Xueying Wang, Shilin Zhou, Yingqian Wang

**Abstract**—Rotated object detection in aerial images has received increasing attention for a wide range of applications. However, it is also a challenging task due to the huge variations of scale, rotation, angle, aspect ratio of objects, and densely arranged targets. Most existing methods heavily rely on amounts of pre-defined anchors with different scales, angles, and aspect ratios, and are optimized with a distance loss. Therefore, these methods are sensitive to anchor hyper-parameters and easily suffer from the performance decrease caused by boundary discontinuity. To handle this problem, we propose a dense anchor-free rotated object detector named DARDet for aerial image rotated object detection in this paper. Our DARDet directly predicts five parameters of rotated boxes at each foreground pixel of feature maps. We design a new alignment convolution module to extract aligned features and introduce a PLoU loss for precise and stable regression. Our method achieves state-of-the-art performance on three commonly used aerial objects datasets (i.e., DOTA, HRSC2016, and UCAS-AOD) while keeping high efficiency<sup>1</sup>.

**Index Terms**—Rotated object detection, Aerial images, Deep convolution neural networks, Anchor-free detector.

## I. INTRODUCTION

OBJECT detection in aerial images plays an important role in numerous applications such as intelligent transportation, port management, and urban planning [1]. However, it is also a challenging task since the objects typically have different scales and aspect ratios. Moreover, the objects are usually displayed in arbitrary directions and are densely packed. The oriented bounding box (OBB) can compactly enclose the target object and is widely used for aerial image object detection.

Most existing rotated object detection methods are anchor-based frameworks. They introduce an additional angle dimension based on general rectangular bounding box object detection and use simple distance loss. For instance, R<sup>2</sup>PN [2] proposes rotated region proposal networks to generate multi-orientated proposals. RoI-Trans [3] learns the spatial transformation from horizontal bounding boxes (HBB) to OBB. In S<sup>2</sup>A-Net [4], an anchor refinement network is proposed to generate high-quality anchors, and only one square anchor is used at each pixel of feature maps. SCRDet [5] applies an IoU-smooth L1 loss to smooth the loss near the boundary. CSL [6] and DCL [7] convert angle regression into a classification task

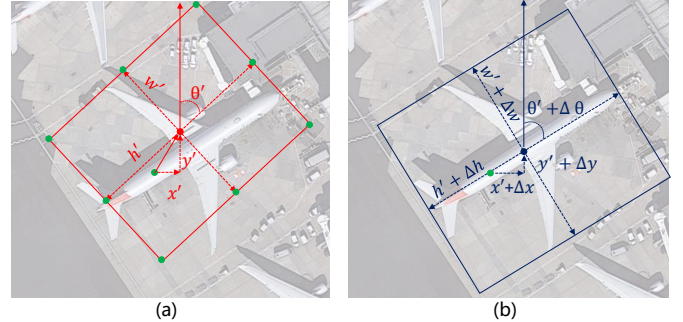


Fig. 1: DARDet directly predicts a 5D vector  $(x', y', w', h', \theta')$  to encode the OBB at each foreground pixel. (a) shows the initial stage. The green dots illustrate the sampling locations of alignment convolution module. We use these locations as offset field and fed them into deformable convolutions to extract aligned features. (b) shows the refining stage, we predict a 5D offset vector  $(\Delta x, \Delta y, \Delta w, \Delta h, \Delta \theta)$  based on the aligned feature. The initially regressed OBB (in red) is then refined into a more accurate one (in blue) by adding the offset.

to handle the boundary problem. Although these methods have achieved promising performance, these models are sensitive to anchor hyper-parameters or easily suffer from the performance decrease caused by boundary discontinuity issue (see Section II-A).

Anchor-free detectors can avoid hyperparameters related to anchor boxes by eliminating the predefined anchor boxes. At present, several anchor-free rotated detectors are proposed. For example, O<sup>2</sup>D-Net [8] detects oriented objects by predicting a pair of middle lines. BBAVectors [9] uses the box boundary-aware vectors to present OBB. CHPDet [10] presents ships as head, center points, and the corresponding size. PolarDet [11] represents the oriented objects by using polar coordinate. However, these methods present rotated boxes in a complex form to solve the boundary discontinuity issue. The features used in these methods are not aligned with the rotated boxes. Consequently, their performance lags behind that of the anchor-base detectors.

In this paper, we propose a dense anchor-free rotated object detector on the VarifocalNet [12] architecture. VarifocalNet is a dense object detector based on FCOS [13]. It learns IoU-aware classification scores (IACS) that simultaneously represent the classification confidence and localization accuracy. In this paper, we extend the VarifocalNet to rotated object detection. In regression branch, as shown in Fig. 1, instead of predicting a 4D vector  $(l, t, r, b)$ , we directly predict a

This work was partially supported in part by the National Natural Science Foundation of China (Nos. 61903373, 61401474, 61921001).

Feng Zhang, Xueying Wang, Shilin Zhou, Yingqian Wang are with the College of Electronic Science and Technology, National University of Defense Technology (NUDT), P. R. China. Emails: {zhangfeng01, wangxueying, slzhou, wangyingqian16}@nudt.edu.cn. (Corresponding author: Xueying Wang)

<sup>1</sup><https://github.com/zf020114/DARDet>

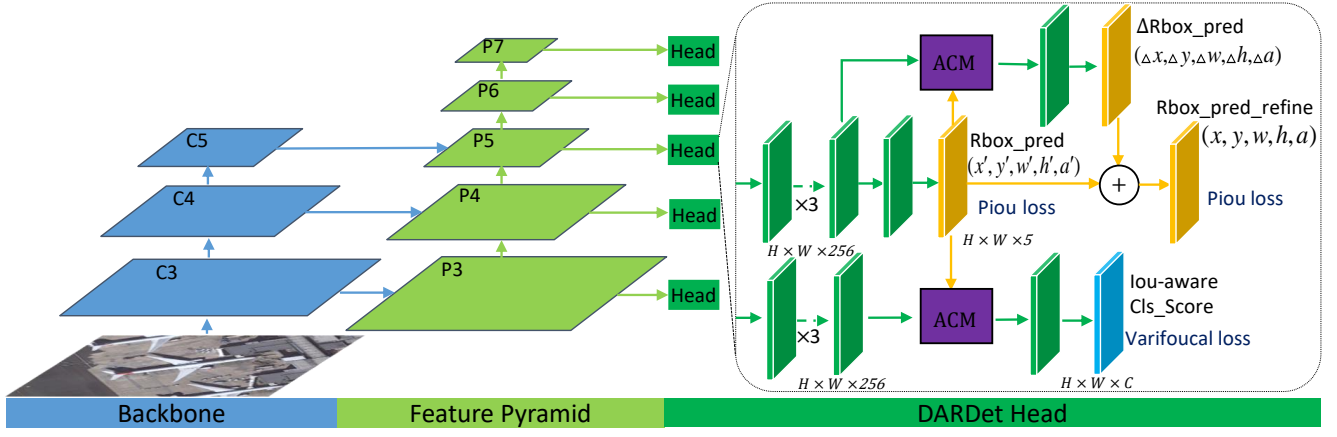


Fig. 2: An overview of our DARDet. Here,  $C3 - C5$  denote the feature maps of the backbone and  $P3 - P7$  denote the feature pyramid. Convolution layers in the same color have the same channels in DARDet head. The DARDet head contains two sub-networks. The localization subnet consists of two stages. In the initial stage, we directly predict five parameters of OBB and get aligned features by alignment convolution module (ACM). In the refining stage, we refine the OBB by predicting a 5D offset vector. The other classification sub-network is similar to the refining stage of the localization sub-network and is used to predict the IACS optimized by Varifocal loss.

5D vector  $(x', y', w', h', \theta')$  encoding the OBB. An efficient alignment convolution module is designed to align features with OBB. We also introduce a Piou loss [14] to handle the boundary discontinuity problem.

The main contributions of this paper are summarized as follows.

- We develop a new dense anchor-free rotated object detection architecture (DARDet), which directly predicts five parameters of OBB at each spatial location.
- We design a new efficient alignment convolution module to extract aligned features, which is used to refine coarse OBB and estimate IACS.
- Our method can effectively handle the boundary discontinuity problem and achieve state-of-the-art performance on the DOTA, UCAS-AOD, and HRSC2016 datasets with high efficiency.

## II. PROPOSED METHOD

In this section, we first explain the boundary discontinuity issue in Section II-A, then introduce our DARDet in Section II-B. In Section II-C, we describe the alignment convolution module. Finally, we introduce the Piou loss in 4.

### A. Boundary discontinuity

The boundary discontinuity refers to the sharp loss increase at the boundary due to the periodicity of the angular and exchangeable property of edges. As shown in Fig. 3, when the angle is near the boundary, it is difficult to judge whether the angle is near  $0^\circ$  or near  $180^\circ$ , because the definition range of angle is  $[0^\circ - 180^\circ)$ . Similarly, when the OBB is approximately square, it is difficult for the model to distinguish which angle to regress, because the definition of angle is the angle between the long edge and the y-axis. When the judgment is wrong, the distance loss will be very large.

### B. Overall Pipeline

As shown in Fig. 2, our DARDet consists of a feature extraction model and DARDet head. The feature extraction model consists of a backbone and FPN. The DARDet head consists of two subnets. The localization subnet takes the feature map from each level of the feature pyramid as its input and applies three  $3 \times 3$  Conv layers to produce a feature map with 256 channels. It performs OBB regression and subsequent refinement in the initial stage and the refining stage, respectively. In the initial stage, the subnet applies a Conv layer to generate a 5D vector  $(x', y', w', h', \theta')$  per spatial location and extract aligned features by alignment convolution module. In the refining stage, the subnet convolves the aligned feature map to produce 5D offsets vector  $(\Delta x, \Delta y, \Delta w, \Delta h, \Delta \theta)$  which is added by the initial OBB location vector for generating the refined OBB  $(x, y, w, h, \theta)$ . Piou loss is used to optimize the OBB. The other classification subnet is used for estimating the IACS, which has a similar architecture with the refining stage of the localization subnet. It outputs IACS, a vector with  $C$  (the class number) channels, which represents both the classification confidence and localization accuracy. We used Varifocal loss for training the dense rotated object detector to predict the IACS.

### C. Alignment Convolution Module

We design a new efficient alignment convolution module which aligns features with OBB by using the representation with deformable convolution. Specifically, given a sampling location  $(i, j)$  on feature map, we first regress an initial OBB vector  $(x', y', w', h', \theta')$ . With this initial OBB, we heuristically select nine sampling points which are the four vertices, the midpoints of the four edges of the OBB, and the sampling location. As shown in Fig. 1(a), the green dots in the figure represent nine sampling localizations. These nine localizations are then mapped onto the feature map

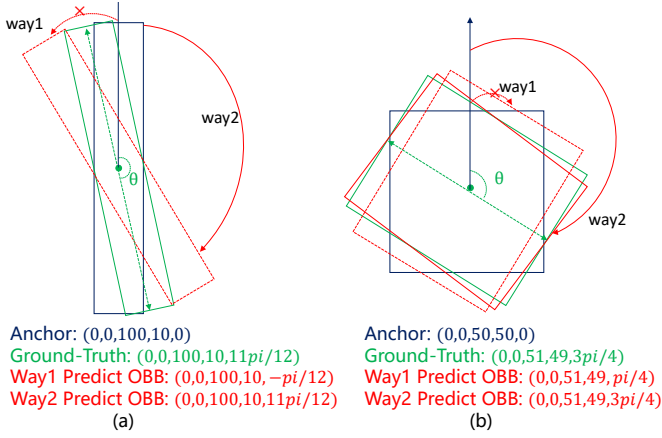


Fig. 3: An illustration of boundary discontinuity problem. The solid and the dotted arrow indicate the correct and wrong regression way, respectively. (a) illustrates the boundary discontinuity due to the periodicity of angles. (b) shows boundary discontinuity caused by the changeable edges.

and features at the projecting points are convolved by the deformable convolution to extract aligned features. This new module is computationally efficient because these points are manually selected with negligible calculation.

#### D. PIoU Loss

Pixels-IoU (PIoU) loss can directly reflect the IoU and its local optimum. This loss is proposed to jointly correlate the five parameters of OBB for checking the position (inside or outside the intersection) and the contribution of each pixel. The intersection areas is calculated by the number of interior pixels. As shown in Fig. 4 (a), for a given OBB  $b$  (blue rotated box encoded by  $(x, y, w, h, \theta)$ ) and a pixel  $p_{i,j}$  in image, to judge the relative location between  $p_{i,j}$  and  $b$ , we define the binary function as follows:

$$\delta(p_{i,j} | b) = \begin{cases} 1, & d_{i,j}^w \leq \frac{w}{2}, d_{i,j}^h \leq \frac{h}{2} \\ 0, & \text{otherwise} \end{cases}, \quad (1)$$

where  $d_{i,j}^h$  and  $d_{i,j}^w$  denote the distance along horizontal and vertical direction, respectively. Since Eq. 1 is not continuous and differentiable, we approximate this binary function by multiplying the following two kernels:

$$F(p_{i,j} | b) = K(d_{i,j}^w, w) K(d_{i,j}^h, h). \quad (2)$$

The kernel function  $K(d, s)$  is defined by

$$K(d, s) = 1 - \frac{1}{1 + e^{-k(d-s)}}. \quad (3)$$

The intersection area  $S_{b \cup b'}$  and union area  $S_{b \cup b'}$  between  $b$  and  $b'$  are approximated by:

$$S_{b \cap b'} \approx \sum_{p_{i,j} \in B_{b,b'}} F(p_{i,j} | b) F(p_{i,j} | b'), \quad (4)$$

$$S_{b \cup b'} \approx w \times h + w' \times h' - S_{b \cap b'}. \quad (5)$$

Then PIoU is computed as:

$$PIoU(b, b') = \frac{S_{b \cap b'}}{S_{b \cup b'}}. \quad (6)$$

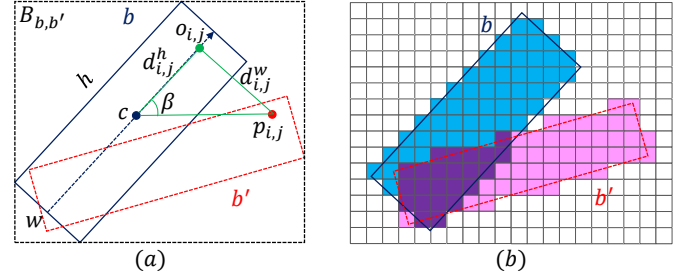


Fig. 4: Schematic diagram of calculating PIoU. (a) Components involved in determining the relative position (inside or outside) between a pixel  $p_{i,j}$  (red point) and an OBB  $b$  (blue rotated box).  $o_{i,j}$  (green point) is the intersection point between the mid-perpendicular line and its orthogonal line to  $p_{i,j}$ . OBB center  $c$  (blue point),  $p_{i,j}$  and  $o_{i,j}$  form a triangle. The length of the two right-angled edges of the triangle are denoted by  $d_{i,j}^h$ ,  $d_{i,j}^w$ . We use a binary function to judge the relative location. (b) Statistics of all pixels to calculate PIoU. The intersection area is in purple, and the area made up of yellow, blue and purple is the union region.

The PIoU loss of the two regression ways in Fig. 3 is almost equal and the loss is continuous under the boundary condition, So it can solve the problem of boundary discontinuity.

### III. EXPERIMENTS

#### A. Datasets

We evaluate our method on the DOTA [15], HRSC2016 [16] and UCAS-AOD [17] datasets.

1) *DOTA*: DOTA is a large dataset for oriented object detection. It contains 2806 images and 15 common object categories. In the ablation study, we use the training set for training and the validation set for evaluation. We cropped the images into  $1024 \times 1024$  patches with a stride of 824 and performed random flipping for data augmentation. For comparison with other methods, all images are cropped with a stride of 512. In multi-scale experiments, we prepared data at three scales (0.5, 1.0, 1.5). Random rotation was used for training and test.

2) *HRSC2016*: The HRSC2016 dataset is a challenging dataset for ship detection in aerial images. The training, validation, and test sets include 436, 181, and 444 images, respectively. The image size of this dataset ranges from  $300 \times 300$  to  $1500 \times 900$ . We used the training and validation sets for training and evaluated the performance on the test set. All images were cropped to patches of size  $1024 \times 1024$ . The aspect ratio was kept unchanged in our experiments.

3) *UCAS-AOD*: The UCAS-AOD dataset contains 1510 aerial images of about  $659 \times 1280$  pixels and 14596 instances of two categories including plane and car. We randomly sampled 1132 images for training and 378 images for testing. All images were cropped into patches of size  $672 \times 672$ .

#### B. Implementation Details

In the ablation study, we used ResNet50 as backbone and trained the models for 12 epochs on a single 11GB NVIDIA



TABLE I: Ablation study for DARDet on the DOTA dataset. We use the training set for training and validation set for evaluation.

	Different Settings of DARDet		
Alignment Convolution Module		✓	✓
PloU loss			✓
mAP	63.19	66.98	72.44

RTX 2080Ti GPU with a batch size of 6. For performance comparison with other state-of-the-art detectors, ResNet50 and Res2Net101 was used as the backbone and trained the models for 24 epochs. We also expanded the receptive field of our detector by replacing ordinary convolutions with deformable convolutions at the last stage of the backbone. In all experiments, SGD optimizer was adopted with an initial learning rate of 0.01 and the learning rate was divided by 10 at each decay step. The momentum and weight decay were 0.9 and 0.0001, respectively. We also adopted linear warmup for 500 iterations. The speed of the network was measured on NVIDIA RTX 2080Ti GPU with a batch size of 1 and  $1024 \times 1024$  image size. The result was evaluated under VOC2007 metrics.

### C. Ablation Study

In this subsection, we present ablation experiments to investigate our models.

1) *Modified VarifocalNet as baseline*: We extend the VarifocalNet architecture to the oriented object detection task by directly predicting a 5D vector to encode the OBB and using smooth L1 loss in the regression branch. As shown in Table I, Modified VarifocalNet achieves an mAP of 63.19%, which demonstrates that our baseline achieves competitive performance.

2) *Effectiveness of Alignment Convolution Module*: We use the alignment convolution module to extract aligned features and then simultaneously estimate the IACS and refine the OBB. As shown in Table I, compared with the baseline, our method improves mAP by 3.8% to 66.98%.

3) *Effectiveness of PloU loss*: When we replace the smooth L1 loss with PloU loss, the overall mAP is improved from 66.98% to 72.44%. It demonstrates the effectiveness of the PloU loss.

Qualitative detection results of the baseline method and our DARDet are visualized in Fig. 5 (a). As shown in the first three columns of the figure, almost all the objects are detected correctly. These results suggest that our method can handle challenging situations even with a high aspect ratio or very cluttered scenes. In the last two columns, the main objects in the images are baseball diamonds and planes. Some objects are almost equal in length and width, so the use of distance loss will encounter boundary problems. Specifically, it is difficult for the model to distinguish which side is longer. To reduce loss, the model regresses the angle to the intermediate value of the two possible correct angles. Therefore, we can see that almost all the baseball diamonds and some planes in the baseline results deviate from the ground truth by  $45^\circ$ . However, our method can solve the boundary problems effectively and obtains accurate results.

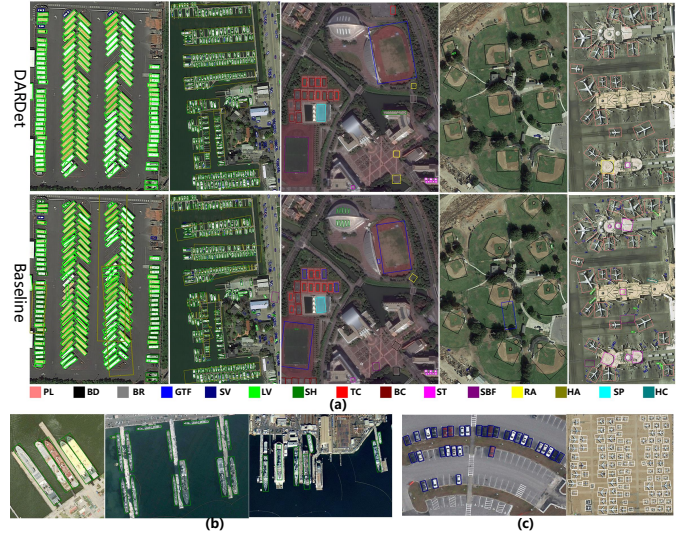


Fig. 5: Some detection results on three commonly used aerial objects datasets. (a) is qualitative comparisons between the proposed DARDet and the baseline method on the DOTA dataset. (b) and (c) are the sample results achieved by DARDet on the HRSC2016 and UCAS-AOD datasets, respectively. Our method achieves promising results when detecting dense, rotated, clutter, and high aspect ratio objects.

### D. Comparison with the SOTA Methods

1) *Results on DOTA*: We compare our DARDet with other state-of-the-art detectors for OBB task. As shown in Table II, our method achieves the state-of-the-art detection accuracies (79.57% in mAP) with multi-scale training and testing. Our single-scale DARDet with ResNet50 achieves an mAP of 77.61%, which outperforms all single-scale models and most multi-scale models without bells and whistles. In terms of the inference speed, our DARDet is much faster than the other methods except S<sup>2</sup>A-Net, but the mAP of our method was higher than the fastest algorithm by 3.5 %.

2) *Results on HRSC2016*: The HRSC2016 contains lots of thin and long ships with arbitrary orientations. The comparative results between our method and some state-of-the-art methods on the HRSC2016 dataset are shown in Table III. Our method achieves the best performance overall the compared methods, at an accuracy of 90.37%.

3) *Results on UCAS-AOD*: Our method also achieves the best performance on the UCAS-AOD dataset, with an mAP of 90.37%. Table IV shows the comparative results.

## IV. CONCLUSION

In this paper, we proposed a new dense anchor-free detection framework to detect rotated objects in aerial images. Our method detects objects by directly predicting a 5D vector at each foreground pixel. Moreover, an alignment convolution module is designed for extracted aligned convolutional features, and PloU loss is introduced to optimize the OBB. All of these designs introduce significant performance improvements. Experimental results on the DOTA, HRSC2016, and UCAS-AOD datasets demonstrate that our method achieves state-of-

TABLE II: Comparisons with state-of-art methods for the DOTA OBB task. MS indicates multi-scale training and testing. The results in red and blue indicate the best and second-best results of each column.

Method	Backbone	MS	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP	FPS
ROI-Trans [3]	ResNet101		88.64	78.52	43.44	75.92	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56	9.4
SCRDet [5]	ResNet50		89.48	79.83	50.43	62.51	76.72	74.31	85.42	<b>90.86</b>	84.19	82.51	42.84	60.15	74.18	70.62	51.81	71.72	4.4
CSL [6]	ResNet50		89.09	79.22	45.78	67.94	66.58	56.29	71.55	90.80	80.50	76.61	58.53	61.01	59.92	69.51	57.59	68.73	5.1
R <sup>3</sup> Det-DCL [7]	ResNet50		89.79	77.80	46.84	65.84	74.87	74.96	85.70	90.23	79.32	84.06	56.59	63.77	57.72	67.62	54.04	71.21	5.4
O <sup>2</sup> -DNet [8]	Hourglass104		89.31	82.14	47.33	61.21	71.32	74.03	78.62	90.76	82.23	81.36	60.93	60.17	58.21	66.98	64.03	71.04	—
BBAVectors [9]	ResNet101		88.35	79.96	50.69	62.18	78.43	78.98	87.94	90.85	83.58	84.35	54.13	60.24	65.22	64.28	55.70	72.32	6.9
PolarDet [11]	ResNet50		89.73	87.05	45.30	63.32	78.44	76.65	87.13	90.79	80.58	85.89	60.97	67.94	68.20	74.63	68.67	75.02	—
S <sup>2</sup> A-Net [4]	ResNet50		89.11	82.84	48.37	71.11	78.11	78.39	87.25	90.83	84.90	85.64	60.36	62.60	65.26	69.13	57.94	74.12	<b>13.5</b>
DARDet (ours)	ResNet50		88.89	84.31	55.32	75.49	<b>80.33</b>	81.69	88.24	<b>90.88</b>	83.62	<b>87.46</b>	59.85	65.60	76.86	<b>80.46</b>	65.17	77.61	<b>12.6</b>
SCRDet [5]	ResNet101	✓	<b>90.05</b>	84.39	55.44	73.99	77.54	71.11	86.05	90.67	<b>87.32</b>	87.08	69.62	<b>68.90</b>	73.74	71.29	65.08	76.81	—
CSL [6]	ResNet152	✓	<b>90.25</b>	<b>88.53</b>	54.64	75.31	70.44	73.51	77.62	90.84	86.15	86.69	69.60	68.04	73.83	71.10	68.93	76.17	—
R <sup>3</sup> Det-DCL [7]	ResNet152	✓	89.26	83.60	53.54	72.76	79.04	82.56	87.31	90.67	86.59	86.98	67.49	66.88	73.29	70.56	<b>69.99</b>	77.37	—
PolarDet [11]	ResNet101	✓	89.65	<b>87.07</b>	48.14	70.97	78.53	80.34	87.45	90.76	85.63	86.87	61.64	70.32	71.92	73.09	67.15	76.64	—
S <sup>2</sup> A-Net [4]	ResNet50	✓	88.89	83.60	<b>57.74</b>	<b>81.95</b>	79.94	<b>83.19</b>	<b>89.11</b>	90.78	84.87	<b>87.81</b>	<b>70.30</b>	<b>68.25</b>	<b>78.30</b>	77.01	<b>69.58</b>	<b>79.42</b>	—
DARDet (ours)	Res2Net101	✓	87.66	85.02	<b>58.51</b>	<b>82.32</b>	<b>80.56</b>	<b>85.54</b>	<b>88.85</b>	90.84	<b>86.39</b>	87.12	<b>73.08</b>	63.32	<b>77.92</b>	<b>77.61</b>	68.73	<b>79.57</b>	—

TABLE III: Detection accuracy on the HRSC2016 dataset.

Method	Backbone	mAP <sub>0.5</sub>
R <sup>2</sup> CNN [18]	Resnet101	73.07
ROI-trans [3]	Resnet101	86.20
R <sup>3</sup> Det [19]	Resnet101	89.33
BBAVectors [9]	Resnet101	88.60
S <sup>2</sup> A-Net [4]	Resnet101	90.17
FR-Est [20]	Resnet101	89.7
DARDet(ours)	Resnet50	<b>90.37</b>

TABLE IV: Detection accuracy on the UCAS-AOD dataset.

Method	Backbone	car	airplane	mAP <sub>0.5</sub>
FR-O [15]	Resnet101	86.87	89.86	88.36
ROI-trans [3]	Resnet101	87.99	89.90	88.95
FPN-CSL [6]	Resnet101	88.09	90.38	89.23
R <sup>3</sup> Det-DCL [7]	Resnet101	88.15	90.57	89.36
DAL [21]	Resnet101	89.25	90.49	89.87
DARDet(ours)	Resnet50	89.97	90.76	<b>90.37</b>

the-art performance with high efficiency as compared to other detectors.

## REFERENCES

- [1] S. He, H. Zou, Y. Wang, R. Li, and F. Cheng, “Shipsrdet: An end-to-end remote sensing ship detector using super-resolved feature representation,” in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2021.
- [2] Z. Zhang, W. Guo, S. Zhu, and W. Yu, “Toward arbitrary-oriented ship detection with rotated region proposal and discrimination networks,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 11, pp. 1745–1749, 2018.
- [3] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, “Learning roi transformer for oriented object detection in aerial images,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [4] J. Han, J. Ding, J. Li, and G.-S. Xia, “Align deep features for oriented object detection,” *IEEE Transactions on Geoscience and Remote Sensing*, 2021.
- [5] X. Yang, J. Yang, J. Yan, Y. Zhang, T. Zhang, Z. Guo, X. Sun, and K. Fu, “ScrDet: Towards more robust detection for small, cluttered and rotated objects,” in *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*. IEEE, 2019, pp. 8231–8240.
- [6] X. Yang and J. Yan, “Arbitrary-oriented object detection with circular smooth label,” in *European Conference on Computer Vision*. Springer, 2020, pp. 677–694.
- [7] X. Yang, L. Hou, Y. Zhou, W. Wang, and J. Yan, “Dense label encoding for boundary discontinuity free rotation detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15819–15829.
- [8] H. Wei, Y. Zhang, Z. Chang, H. Li, H. Wang, and X. Sun, “Oriented objects as pairs of middle lines,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 268–279, 2020.
- [9] J. Yi, P. Wu, B. Liu, Q. Huang, H. Qu, and D. Metaxas, “Oriented object detection in aerial images with box boundary-aware vectors,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 2150–2159.
- [10] F. Zhang, X. Wang, S. Zhou, Y. Wang, and Y. Hou, “Arbitrary-oriented ship detection through center-head point extraction,” *arXiv preprint arXiv:2101.11189*, 2021.
- [11] P. Zhao, Z. Qu, Y. Bu, W. Tan, and Q. Guan, “Polardet: A fast, more precise detector for rotated target in aerial images,” *International Journal of Remote Sensing*, vol. 42, no. 15, pp. 5821–5851, 2021.
- [12] H. Zhang, Y. Wang, F. Dayoub, and N. Sunderhauf, “Varifocalnet: An iou-aware dense object detector,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8514–8523.
- [13] Z. Tian, C. Shen, H. Chen, and T. He, “FCOS: fully convolutional one-stage object detection,” in *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*. IEEE, 2019, pp. 9626–9635.
- [14] Z. Chen, K. Chen, W. Lin, J. See, H. Yu, Y. Ke, and C. Yang, “Piou loss: Towards accurate oriented object detection in complex environments,” in *European Conference on Computer Vision*. Springer, 2020, pp. 195–211.
- [15] G. Xia, X. Bai, J. Ding, Z. Zhu, S. J. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, “DOTA: A large-scale dataset for object detection in aerial images,” in *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*. IEEE Computer Society, 2018, pp. 3974–3983.
- [16] Z. Liu, L. Yuan, L. Weng, and Y. Yang, “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. SCITEPRESS, 2017, pp. 324–331.
- [17] C. Li, C. Xu, Z. Cui, D. Wang, T. Zhang, and J. Yang, “Feature-attentioned object detection in remote sensing imagery,” in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 3886–3890.
- [18] Y. Jiang, X. Zhu, X. Wang, S. Yang, W. Li, H. Wang, P. Fu, and Z. Luo, “R2cnn: Rotational region cnn for orientation robust scene text detection,” *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018.
- [19] X. Yang, J. Yan, Z. Feng, and T. He, “R3det: Refined single-stage detector with feature refinement for rotating object,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 4, 2021, pp. 3163–3171.
- [20] K. Fu, Z. Chang, Y. Zhang, and X. Sun, “Point-based estimator for arbitrary-oriented object detection in aerial images,” *IEEE Transactions on Geoscience and Remote Sensing*, 2020.
- [21] Q. Ming, Z. Zhou, L. Miao, H. Zhang, and L. Li, “Dynamic anchor learning for arbitrary-oriented object detection,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 3, 2021, pp. 2355–2363.