

Optimized Point Set Representation for Oriented Object Detection in Remote-Sensing Images

Junjie Song, Lingjuan Miao¹, Zhiqiang Zhou¹, *Member, IEEE*, Qi Ming¹, and Yunpeng Dong

Abstract—How to represent the object more appropriately in oriented object detection is an essential problem to be solved, because there are many solutions for the object represented. It is a relatively novel approach to represent objects as a number of sample points useful for both localization and recognition. However, the current point-set-based representation methods do not effectively supervise all points for learning, and the internal information of the convex hull in the point set cannot be effectively learned. Therefore, this letter proposes point set distance (PSD) loss, which learns set-to-set supervision of objects to effectively represent objects. Besides, most of the current sample selection strategies are based on the Intersection over Union (IoU), but these methods cannot comprehensively measure candidate samples' quality. To select high-quality point sets, we propose to use the probability distribution of point sets to select the positive samples. Our probabilistic point set sample selection (PPSS) scheme effectively exploits the classification information, regression information, and distribution characteristics of the point set. Experimental results on remote-sensing image datasets, including DOTA, DIOR-R, and HRSC2016, demonstrate the proposed method for arbitrary-oriented object detection achieves consistent and substantial improvements.

Index Terms—Oriented object detection, points representation, remote-sensing images, sample selection.

I. INTRODUCTION

OBJECT detection in remote-sensing images is to accurately locate and identify the objects of interest. It has become crucial in many real-world applications such as town planning, strategic deployment in the military field, and Earth observation [1]. Object detection in remote-sensing images has very promising applications, but there are still some challenges that need to be overcome.

Due to the arbitrary orientation, dense distribution, and complex background of the object in remote-sensing images, the horizontal bounding boxes do not represent their semantic and localization information properly, resulting in the performance of various detectors being greatly limited. As a result, the currently popular approaches of object detection in remote-sensing images use oriented bounding boxes to represent the objects, for which most angle-based detectors achieve excellent performance. However, the boundary discontinuity and square-like problem are obstacles to high-precision

Manuscript received 31 May 2023; revised 20 August 2023; accepted 6 September 2023. Date of publication 12 September 2023; date of current version 2 October 2023. (*Corresponding author: Zhiqiang Zhou.*)

The authors are with the School of Automation, Beijing Institute of Technology, Beijing 100081, China (e-mail: bitsongjj@gmail.com; miaolingjuan@bit.edu.cn; zhzhzhou@bit.edu.cn; chaser.ming@gmail.com; bitdyp@gmail.com).

Digital Object Identifier 10.1109/LGRS.2023.3314517

1558-0571 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See <https://www.ieee.org/publications/rights/index.html> for more information.

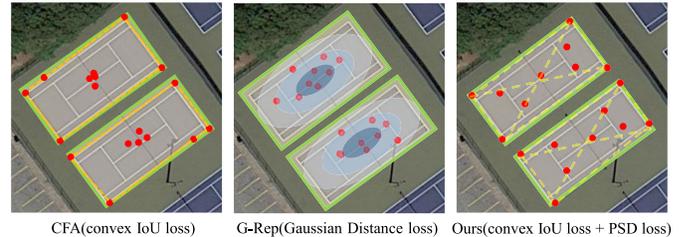


Fig. 1. Illustration of point representation for oriented object detection. CFA [4] cannot effectively capture information inside the object, and G-Rep [7] does not capture the boundary information of the object. Our method makes the points better distributed on the object.

locating, as detailed in [2] and [3]. In addition, features extracted from the oriented bounding box are heavily influenced by background information or uninformative foreground areas that contain little semantic information. Many researchers [4], [5], [6], [7] try to abandon the anchor and use a number of points to represent the object to solve the above problems. This point set learning method uses the adaptive points representation for oriented object detection, which exploits the captured geometric and semantic information of arbitrarily oriented objects. As shown in Fig. 1, CFA [4] cannot effectively capture the semantic information of objects. It uses convex hull Intersection over Union (IoU) loss to optimize the position information of the point set. However, the convex hull IoU loss in the training stage only supervises the geometric information of the point set, and its internal points are not optimized, resulting in redundant information of the point set that cannot well capture semantic information. G-Rep [7] converts the point set into Gaussian distribution and uses the Gaussian metric-based regression loss to optimize the position of the point set. However, the Gaussian metric-based regression loss guides the point set to distribute in the middle region of the object and does not capture the boundary information of the object.

Besides, the most popular strategy to determine positive samples is to use the IoU metric between anchors and ground truth (GT). However, the strategy has an obvious limitation in that it ignores the actual content of the intersecting regions, and these regions may contain noisy backgrounds, other objects, or regions that are not meaningful for detection. Several studies [8], [9] have pointed out this limitation and proposed various new sample selection strategies to improve the performance of detectors. However, due to the complex background, dense distribution, and arbitrary orientation of

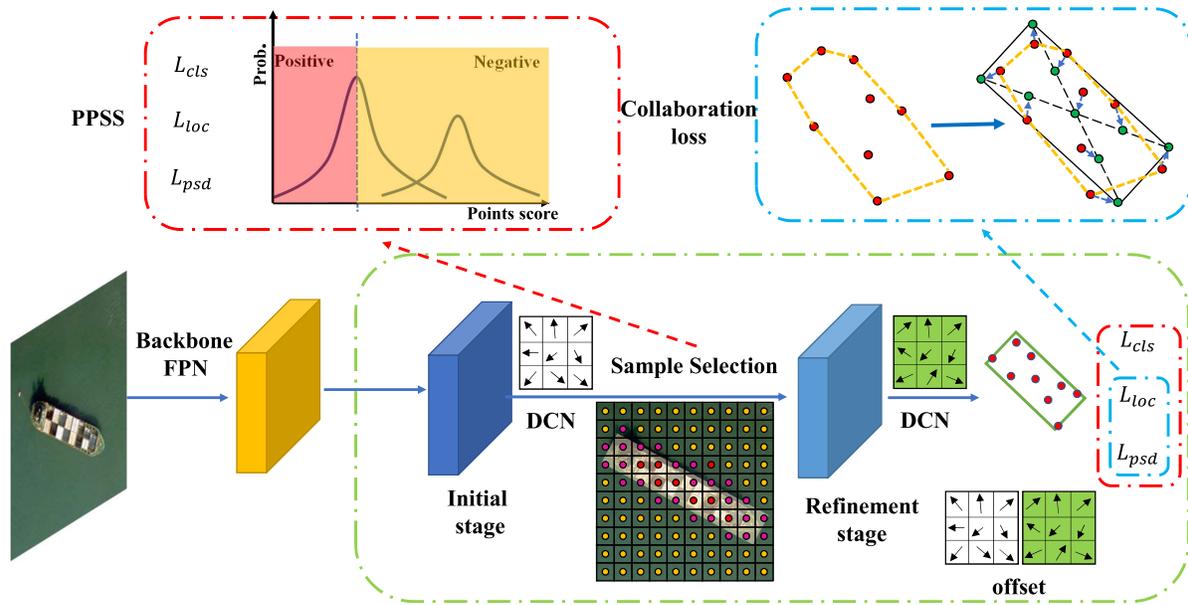


Fig. 2. Architecture of the proposed method. PPSS is used to select more promising samples and prevent filtering out high-quality samples. The collaboration loss is composed of IoU loss and our proposed PSD loss, which achieves better convergence and performance. DCN is deformable convolution.

objects in remote-sensing images, the limitation still cannot be effectively eliminated.

In this letter, to solve the above issues, we propose the point set distance (PSD) loss, which provides set-to-set supervision to better capture the geometric and semantic information of the objects. We also proposed a probabilistic point set sample selection (PPSS) strategy for the point-based model, which combines regression scores, classification scores, and point set similarity to select high-quality positive samples.

II. PROPOSED METHOD

A. Overview of the Proposed Method

The proposed method is built on RepPoints [10], which is a representative-oriented anchor-free method. As shown in Fig. 2, the whole method consists of a backbone network with FPN, initial detection head, and refinement detection head. The initial detection head generates coarse point sets, which are converted into convex hulls using the Jarvis March algorithm [11] and further refined in the refinement detection head inspired by [4]. The proposed PSD loss assists IoU loss in the refinement stage to provide set-to-set supervision of objects to adaptively select suitable matching pairs according to the current optimization situation. The PSD loss guides the point set to characterize the object more efficiently, learning the geometric and semantic information of objects. The proposed PPSS scheme is adopted in the refinement detection head, which helps to consider the classification score, regression score, and points similarity of the point set simultaneously. Then, the candidate samples are adaptively classified into positive samples and negative samples based on the above multitask information, to select high-quality positive samples. PSD loss and the PPSS scheme are described in the following sections.

B. Set-to-Set Supervision for Optimal Object Representation

In point-set-based models, deformable convolution is used to adjust the locations of sampled points with offsets predicted

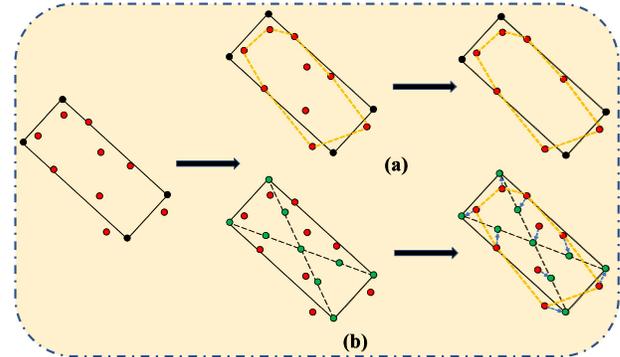


Fig. 3. Illustration of PSD loss and IoU loss optimization process (a) denotes the process of the IoU loss function and (b) denotes the process of our PSD loss assists IoU loss to optimize the point set. The green points are based on GT, and the additional construction points are in the middle of the four corners and the center of the GT. The red points denote the point set predicted by the refinement stage.

by the model. However, current methods use IoU loss or Gaussian distance loss, and these methods cannot effectively supervise the point set and only optimize the location of points based on geometric information. The points inside the convex hull lack additional supervision information for position optimization, which will result in the sampling positions of the deformable convolution not covering the object appropriately. To capture high-quality semantic and geometric information of objects, we propose a point set similarity distance (PSD) loss, which introduces set-to-set supervision to make the point set evenly distributed on the object. PSD loss assists IoU loss in providing guidance information for the optimization of model parameters. The point set uniformly covers the object which is more appropriate to represent its semantic and geometric information.

In Fig. 3, the nine red points denote the point set predicted by the refinement stage. They are then optimized using the supervision information provided by the preset green points. Since the regression of the point set is disordered, we regard

the matching of the point set and GT as a dynamic problem and use the Hungarian matching algorithm to solve the above problem. To find a bipartite matching between GT and prediction, we search for a permutation of the nine points with the lowest cost

$$\bar{\sigma} = \operatorname{argmin}_{\sigma \in \mathfrak{S}_N} \sum_i^N L_{\text{match}}(g_i, p_{\sigma(i)}) \quad (1)$$

where \mathfrak{S}_N denotes all possible permutations of indices $1 \sim N$ for these points ($N = 9$), and $L_{\text{match}}(g_i, p_{\sigma(i)})$ is a pair-wise matching cost between g_i of GT and a prediction with the index $\sigma(i)$. $\bar{\sigma}$ is the optimal assignment obtained by minimizing the overall matching cost. Based on the assignment results, the definition of PSD loss is as follows:

$$L_{\text{PSD}}(p, g) = \frac{1}{N} \sum_{i=1}^N L_{\text{reg}}(p_{\bar{\sigma}(i)}, g_i) \quad (2)$$

where $p_{\bar{\sigma}}$ is the optimal bipartite matching and L_{reg} denotes the smooth $L1$ loss.

PSD loss guides point set to well capture semantic information and geometric information of the object. As shown in Fig. 3(b), PSD loss adaptively selects suitable matching pairs according to the current optimization situation and can be collaborated with the IoU loss to further optimize the point set. As a result, as shown in Fig. 1, our method can generate the point set with the distribution more appropriate for representing objects.

C. Probabilistic Point Set Sample Selection

The main-stream label assignment strategies are typically performed based on IoU, but these strategies ignore the actual content of the intersecting regions. Moreover, the RepPoints method lacks direct supervision of the point set, while learning high-quality point sets is crucial to capturing the geometric and semantic features of dense and arbitrarily oriented objects. Therefore, to measure the quality of candidate samples more appropriately, we consider the regression score, classification score, and point set similarity as metrics and combine these three metrics to select high-quality positive samples.

The score of the points set is defined as follows:

$$S = S_{\text{cls}} + S_{\text{loc}} + S_{\text{sim}} \quad (3)$$

where S_{cls} and S_{loc} denote the classification score and regression score of the point set, respectively, and S_{sim} denotes the point set similarity score formulated by the proposed PSD loss. Note that the lower values of S indicate the higher quality of the samples. Let G^{cls} denote the class label for GT, and P^{cls} denote the predicted class confidence based on the learned point set. S_{cls} is defined as

$$S_{\text{cls}} = L_{\text{cls}}(P^{\text{cls}}, G^{\text{cls}}) \quad (4)$$

where L_{cls} denotes focal loss. In (3), we use the loss function S_{loc} to eliminate the gap between classification and regression, which is defined as follows:

$$S_{\text{loc}} = \frac{1}{M} \sum_{n=1}^M L_{\text{loc}}(P_n^{\text{loc}}, G^{\text{loc}}) \quad (5)$$

where M is the number of candidate point sets and L_{loc} is the IoU loss for the oriented polygon. P^{loc} and G^{loc} represent the position of the predicted point set in the refinement stage and the GT box, respectively.

Note that the position of the point set represents the sampling position of the deformable convolution, which indicates the potential of capturing the geometric and semantic information of objects. To fully exploit the information of the point set, S_{sim} is proposed to measure point set similarity as follows:

$$S_{\text{sim}} = L_{\text{PSD}}(P^{\text{loc}}, \mathcal{R}^{\text{loc}}) \quad (6)$$

where \mathcal{R}^{loc} represents the supervised information of the point set that we construct based on GT. L_{PSD} is the proposed PSD loss function.

Based on the obtained point set scores, we use the Gaussian mixture model (GMM) to simulate the score distribution of point sets

$$P(s|\theta) = w_1 \mathcal{N}_1(s; m_1, p_1) + w_2 \mathcal{N}_2(s; m_2, p_2) \quad (7)$$

where \mathcal{N}_1 and \mathcal{N}_2 denote the Gaussian distribution probability densities of positive and negative samples (with the mean and precision parameters (m_1, p_1) and (m_2, p_2) , respectively). w_1 and w_2 are their weights, and s and θ are the point set scores and GMM parameters. Given a set of point set scores, the likelihood of this GMM can be optimized using the expectation-maximization (EM) algorithm. With the parameters estimated by EM, the probability of the point set being a positive or a negative sample can be determined.

III. EXPERIMENTS

A. Datasets

In the experiments, the proposed method is validated on DOTA [12], DIOR-R [13], and HRSC2016 [14] datasets.

1) *DOTA*: DOTA is a large dataset for oriented object detection, which contains 2806 images and 15 categories. We cropped the images into 1024×1024 patches with a stride of 824. We only use a single scale for training and testing.

2) *DIOR-R*: DIOR-R annotated the oriented bounding boxes for every instance, which includes 23 463 images with size 800×800 and 20 categories. The training, validation, and test sets include 5862, 5863, and 11 738 images, respectively. We used the training and validation sets for training and evaluated the performance on the test set.

3) *HRSC2016*: HRSC2016 dataset contains 1061 images with 2886 samples. The image sizes range from 300×300 to 1500×900 . The training, validation, and test sets include 436, 181, and 444 images, respectively. For this dataset, the shorter sides of the images are resized to 800 while the longer sides are kept less than or equal to 1333 with the aspect ratios unchanged.

B. Implementation Details

We chose ResNet50 with FPN as the backbone network for ablation experiments. Our model is built and trained based on the MMRotate [15] platform. We train the model with 40, 40, and 80 epochs for DOTA, DIOR-R, and HRSC2016, respectively. In all experiments, the SGD optimizer is adopted

TABLE I

COMPARISON WITH STATE-OF-THE-ART METHODS ON THE DOTA DATASET. THE RESULTS WITH RED COLOR DENOTE THE BEST RESULTS AND WITH BLUE COLOR PRESENT THE SECOND-BEST RESULTS IN EACH COLUMN

Methods	Backbone	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
Anchor-based																	
FR-O [16]	ResNet50	88.44	73.06	44.86	59.09	73.25	71.49	77.11	90.84	78.94	83.90	48.59	62.95	62.18	64.91	56.18	69.05
DAL [17]	ResNet101	88.61	79.69	46.27	70.37	65.89	76.10	78.53	90.84	79.98	78.41	58.71	62.02	69.23	71.32	60.65	71.78
R ³ Det [18]	ResNet152	89.49	81.17	50.53	66.10	70.92	78.66	78.21	90.81	85.26	84.23	61.81	63.77	68.16	69.83	67.17	73.74
S ² A-Net [19]	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
R ³ Det-DCL [3]	ResNet152	89.78	83.95	52.63	69.70	76.84	81.26	87.30	90.81	84.67	85.27	63.50	64.16	68.96	68.79	65.45	75.54
Oriented R-CNN [20]	ResNet101	88.86	83.48	55.27	76.92	74.27	82.10	87.52	90.90	85.56	85.33	65.51	66.82	74.36	70.15	57.28	76.28
Anchor-free																	
BBAVectors [21]	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	73.32
CFA [4]	ResNet101	89.26	81.72	51.81	67.17	79.99	78.25	84.46	90.77	83.40	85.54	54.86	67.75	73.04	70.24	64.96	75.05
SASM [5]	ResNet50	86.42	78.97	52.47	69.84	77.30	75.99	86.72	90.89	82.63	85.66	60.13	68.25	73.98	72.22	62.37	74.92
G-Rep [7]	ResNet50	87.76	81.29	52.64	70.53	80.34	80.56	87.47	90.74	82.91	85.01	61.48	68.51	67.53	73.02	63.54	75.56
OrientedRepPoint [6]	Swin-T	89.11	82.32	56.71	74.95	80.70	83.73	87.67	90.81	87.11	85.85	63.60	68.60	75.95	73.54	63.76	77.63
Ours	ResNet50	88.99	82.28	54.02	73.32	81.01	81.88	88.13	90.85	86.84	84.59	64.24	67.34	73.56	75.73	56.08	76.59
Ours	Swin-T	88.35	83.79	54.63	75.60	81.43	82.58	88.12	90.87	87.61	86.93	66.85	70.20	75.72	78.63	59.17	78.03

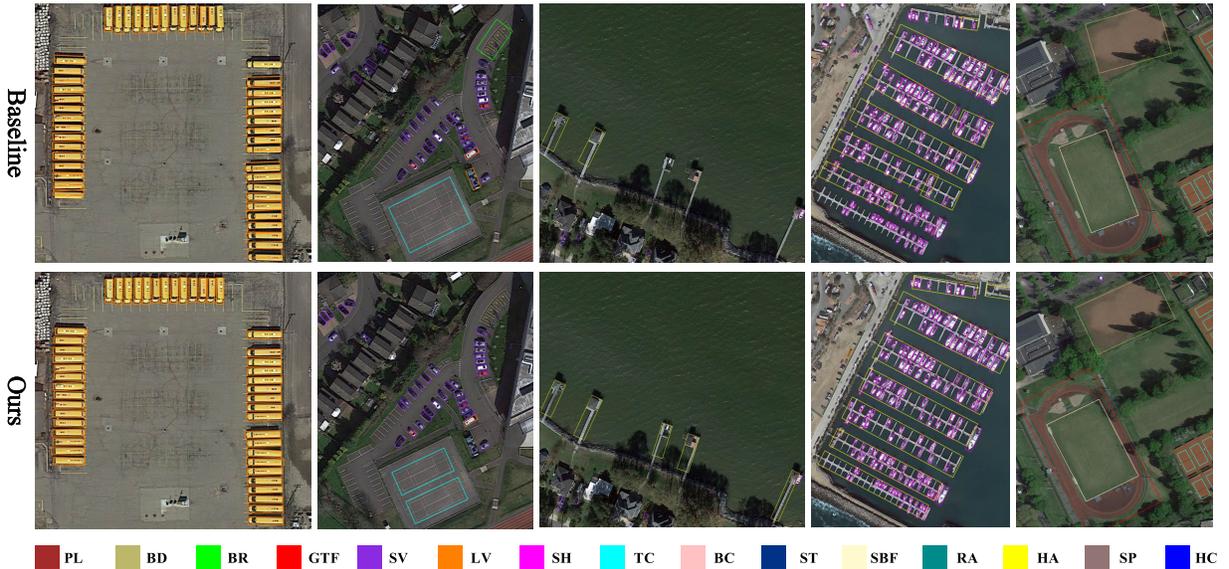


Fig. 4. Comparison of detection results in DOTA with two methods. The upper image is the baseline method while the bottom is our proposed method.

with an initial learning rate of 0.008, and the learning rate is divided by 10 at each decay step. The momentum and weight decay are 0.9 and 0.0001, respectively. We use two RTX 3090 GPUs for training with a total batch size of 4 (two images per GPU). The results are tested on a single RTX 3090 GPU using VOC2007 metrics.

C. Comparisons With State-of-the-Art

Table I shows the comparison results with state-of-the-art methods on the OBB task of the DOTA dataset. Our method with ResNet50 achieves a mAP of 76.59%, which outperforms all single-scale models without bells and whistles. Our methods with Swin Transformer (Swin-T) achieve state-of-the-art detection accuracies (78.03% in mAP) without data augmentation compared with other single-scale methods. The visualization of the baseline and our approach to DOTA are shown in Fig. 4. Compared with Rotate RepPoints, our method achieves higher recall and precision for objects with arbitrary orientation and dense distribution.

The comparison results with state-of-the-art methods on the HRSC2016 dataset are shown in Table II, and our method with

TABLE II

COMPARISON WITH OTHER METHODS ON THE HRSC2016 DATASET

Methods	Backbone	mAP
BBAVectors [21]	ResNet101	88.60
CFA [4]	ResNet50	88.30
R ³ det [18]	ResNet101	89.26
DAL [17]	ResNet50	88.60
R ³ Det-DCL [3]	ResNet101	89.46
OrientedRepPoint [6]	ResNet101	89.30
SASM [5]	ResNet101	90.00
Ours	ResNet50	89.53

ResNet50 achieves competitive performance (89.53% mAP) without any tricks. SASM [5] is good at detecting long-oriented objects, so it can achieve excellent performance in the HRSC2016 dataset.

D. Ablation Studies

1) *Effectiveness of PSD*: As shown in Table III, PSD loss on the DOTA dataset obtains consistent gains on different models. As shown in Table IV, when we introduce the PSD in the regression loss, we obtained gains of 0.23% mAP, 0.33% mAP,

TABLE III
PERFORMANCE IMPROVEMENT OF PSD ON DIFFERENT
REPPPOINTS-BASED MODELS

Methods	CFA	CFA+PSD	O-RP	O-RP+PSD	Ours	Ours+PSD
mAP	74.38	75.13(↑0.75)	75.21	75.92(↑0.71)	76.36	76.59(↑0.23)

TABLE IV
ABLATION STUDY RESULTS BASED ON REPPPOINTS FROM THE DOTA,
DIOR-R, AND HRSC2016 DATASETS. “I” AND “TI” INDICATE THE
INDIVIDUAL IMPROVEMENT AND THE TOTAL
IMPROVEMENT IN MAP

Dataset	$S_{cls}+S_{loc}$	$S_{cls}+S_{loc}+S_{sim}$	PSD	mAP	I	TI
DOTA				68.69		
	✓			75.82	+7.13	+7.13
	✓	✓		76.36	+0.54	+7.67
	✓	✓	✓	76.59	+0.23	+7.90
DIOR-R				55.27		
	✓			69.16	+13.89	+13.89
	✓	✓		70.42	+1.26	+15.15
	✓	✓	✓	70.75	+0.33	+15.48
HRSC2016				75.11		
	✓			88.50	+13.39	+13.39
	✓	✓		89.12	+0.62	+14.01
	✓	✓	✓	89.53	+0.41	+14.42

TABLE V
COMPARISONS WITH DIFFERENT SAMPLE ASSIGNMENT METHODS ON
ROTATED REPPPOINTS DETECTOR

Dataset	Methods				
	Max-IoU [16]	CFA [4]	APAA [6]	SASM [5]	Ours(PPSS)
DOTA	68.69	74.89	75.21	74.92	76.36
DIOR-R	55.27	66.51	66.71	67.41	70.42
HRSC2016	75.11	88.32	88.81	90.00	89.53

0.41% mAP on the DOTA, DIOR-R, and HRSC2016 dataset, respectively.

2) *Effectiveness of PPSS*: We compare the proposed PPSS method with other sample selection schemes for training the rotate RepPoints detector, including Max-IoU [16], CFA [4], APAA [6], and SASM [5]. As shown in Table V, our method achieved 76.36% mAP on the DOTA dataset, which is 7.68% mAP higher than the baseline. Our method achieves great gains on both DIOR-R and HRSC2016 datasets. Our PPSS scheme achieves the best performance on different datasets without complicated operations, which demonstrates that our proposed PPSS is effective for point set learning.

IV. CONCLUSION

In this letter, the PSD loss is proposed to solve the representation ambiguity and accelerate network convergence in oriented object detection. Our method effectively learns the geometric and semantic information of objects, allowing the point set to characterize the object more efficiently. Moreover, to effectively learn the adaptive point set, we introduced the PPSS scheme to select the high-quality points samples for training based on the probability distribution of the point set scores. Extensive experiments on multiple datasets demonstrated the superiority of our approach.

REFERENCES

- [1] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, “Object detection in optical remote sensing images: A survey and a new benchmark,” *ISPRS J. Photogramm. Remote Sens.*, vol. 159, pp. 296–307, Jan. 2020.
- [2] X. Yang, J. Yan, Q. Ming, W. Wang, X. Zhang, and Q. Tian, “Rethinking rotated object detection with Gaussian Wasserstein distance loss,” in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 11830–11841.
- [3] X. Yang, L. Hou, Y. Zhou, W. Wang, and J. Yan, “Dense label encoding for boundary discontinuity free rotation detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15814–15824.
- [4] Z. Guo, C. Liu, X. Zhang, J. Jiao, X. Ji, and Q. Ye, “Beyond bounding-box: Convex-hull feature adaptation for oriented and densely packed object detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8788–8797.
- [5] L. Hou, K. Lu, J. Xue, and Y. Li, “Shape-adaptive selection and measurement for oriented object detection,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 1, Jun. 2022, pp. 923–932.
- [6] W. Li, Y. Chen, K. Hu, and J. Zhu, “Oriented RepPoints for aerial object detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 1819–1828.
- [7] L. Hou, K. Lu, X. Yang, Y. Li, and J. Xue, “G-Rep: Gaussian representation for arbitrary-oriented object detection,” *Remote Sens.*, vol. 15, no. 3, p. 757, Jan. 2023.
- [8] K. Kim and H. S. Lee, “Probabilistic anchor assignment with IoU prediction for object detection,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Glasgow, U.K.: Springer, Aug. 2020, pp. 355–371.
- [9] J. Song, L. Miao, Q. Ming, Z. Zhou, and Y. Dong, “Fine-grained object detection in remote sensing images via adaptive label assignment and refined-balanced feature pyramid network,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 71–82, 2023.
- [10] Z. Yang, S. Liu, H. Hu, L. Wang, and S. Lin, “RepPoints: Point set representation for object detection,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9656–9665.
- [11] R. A. Jarvis, “On the identification of the convex hull of a finite set of points in the plane,” *Inf. Process. Lett.*, vol. 2, no. 1, pp. 18–21, Mar. 1973.
- [12] G.-S. Xia et al., “DOTA: A large-scale dataset for object detection in aerial images,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3974–3983.
- [13] G. Cheng et al., “Anchor-free oriented proposal generator for object detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5625411.
- [14] Z. Liu, L. Yuan, L. Weng, and Y. Yang, “A high resolution optical satellite image dataset for ship recognition and some new baselines,” in *Proc. 6th Int. Conf. Pattern Recognit. Appl. Methods*, 2017, pp. 324–331.
- [15] Y. Zhou et al., “MMRotate: A rotated object detection benchmark using PyTorch,” in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 7331–7334.
- [16] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 91–99.
- [17] Q. Ming, Z. Zhou, L. Miao, H. Zhang, and L. Li, “Dynamic anchor learning for arbitrary-oriented object detection,” in *Proc. AAAI*, vol. 35, no. 3, Jun. 2021, pp. 2355–2363.
- [18] X. Yang, J. Yan, Z. Feng, and T. He, “R3Det: Refined single-stage detector with feature refinement for rotating object,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 4, May 2021, pp. 3163–3171.
- [19] J. Han, J. Ding, J. Li, and G.-S. Xia, “Align deep features for oriented object detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5602511.
- [20] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, “Oriented R-CNN for object detection,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3500–3509.
- [21] 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 *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 2149–2158.