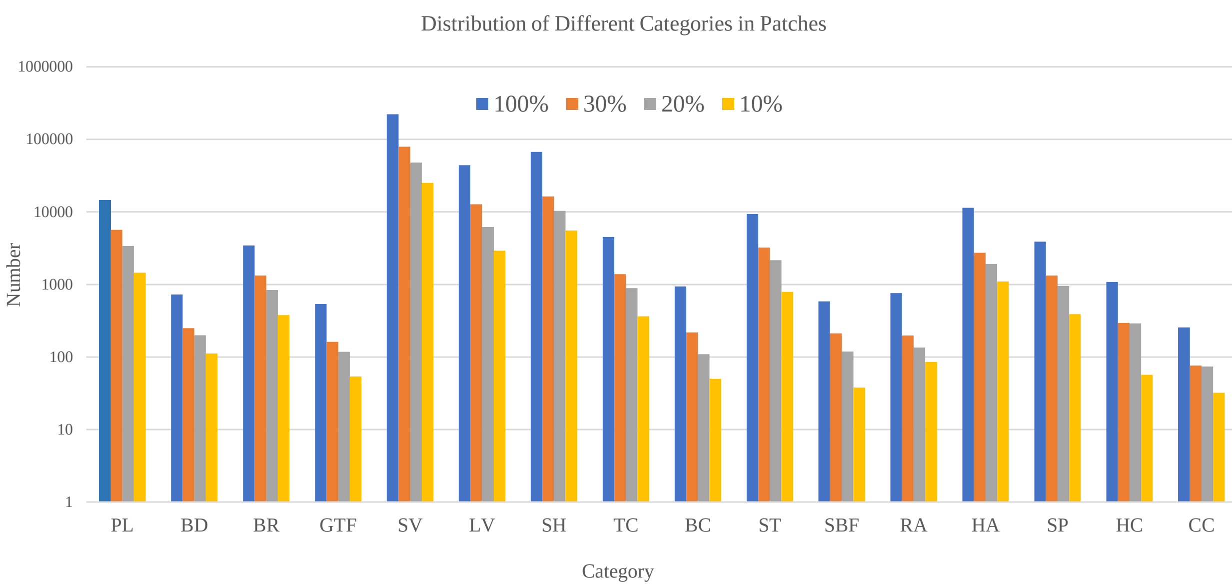
SOOD: Towards Semi-Supervised Oriented Object Detection
(Supplementary Material)

1. Details of Partially Labeled Data Sets

We randomly sample 10%, 20%, and 30% data from DOTA-v1.5-train [5] to form partially labeled data sets. Besides, the 20% set is a subset of the 30% set, and the 10% set is a subset of the 20% set. To maintain the characteristic of the original data, we ensure the partially labeled sets have similar data distributions with DOTA-v1.5-train, as shown in Fig. 1. In this manner, these splits can well reflect the effectiveness of different semi-supervised object detection methods.



Category	Plane	Baseball diamond	Bridge	Ground track field	Small vehicle	Large vehicle	Ship	Tennis court
Abbreviation	PL	BD	BR	GTF	SV	LV	SH	TC
Category	Basketball court	Storage tank	Soccer-ball field	Roundabout	Harbor	Swimming pool	Helicopter	Container crane
Abbreviation	BC	ST	SBF	RA	HA	SP	HC	CC

Figure 1. The top table shows distributions of different categories in the cropped image patches with abbreviations. The bottom table shows the corresponding abbreviation for each category. The distributions of different data splits are similar to the origin (100%). In this case, these splits can well reflect the performance of different semi-supervised object detection methods.

Table 1. Detailed comparison between SOOD and other methods. All experiments are conducted on DOTA-v1.5 under the Fully Labeled Data setting.

Method	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	CC	mAP
FCOS [4]	89.8	76.5	42.1	56.9	54.7	77.9	89.4	90.6	63.0	**70.1**	60.6	66.0	73.5	65.5	56.1	14.7	65.46
Dense Teacher [8]	**89.9**	83.6	**46.5**	50.3	53.6	75.0	88.7	90.7	**69.5**	67.9	65.9	**73.3**	**73.7**	**67.5**	56.9	9.1	66.38
SOOD (ours)	89.7	**84.0**	46.1	**57.9**	**55.4**	**78.6**	**89.6**	**90.8**	66.8	69.0	**69.5**	70.1	**73.7**	66.6	**63.1**	**15.9**	**67.70**

2. Additional Experiments

2.1. Analysis on Result of Each Category

As shown in Tab. 1, we report the detailed results between our SOOD and other methods under the Fully Labeled Data setting. Our SOOD achieves 67.70 mAP, outperforming Dense Teacher [8] for most categories. Specifically, SOOD outperforms Dense Teacher by a large margin for categories named ground track field (GTF), helicopter (HC), and container crane (CC). We think the main reasons are two aspects: 1) the proposed rotation-aware adaptive weighting (RAW) loss can effectively utilize the orientation information, improving performance on orientation-sensitive objects like GTF and CC. 2) the proposed global consistency (GC) loss builds global constraint as an auxiliary, improving performance on dense objects like HC and CC.

Although SOOD outperforms the supervised baseline on basketball court (BC) and roundabout (RA) by a lot, it is lower than Dense Teacher. It is likely that these two categories are often alone and similar to the background, which weakens the effect of GC and leads to sub-optimal supervision. Besides, even though SOOD performs better than Dense Teacher on storage tank (ST), it is worse than the supervised baseline. It may be because that RA is similar to the background and easily confused with other objects, resulting in too much noise in pseudo labels and worse performance.

2.2. SOOD with Anchor-Based Detectors

We additionally adopt our SOOD on anchor-based detectors, e.g., rotated-Faster-RCNN [3] and Oriented R-CNN [6]. As shown in Tab. 2, our SOOD surpasses previous semi-supervised object detection methods on rotated-Faster-RCNN. Besides, on the state-of-the-art method Oriented R-CNN, SOOD still improves the performance by +1.24, which proves our SOOD is also suitable for anchor-based detectors.

Table 2. Comparison of our SOOD and other semi-supervised object detection methods using anchor-based detectors under the Fully Labeled Data setting. All methods are evaluated on DOTA-v1.5-val. * indicates implementation towards oriented objects.

Detector	Method	Publication	mAP	Δ
Faster R-CNN* [3]	Supervised	NeurIPS 2016	66.12	-
	Unbiased Teacher [2]	ICLR 2021	64.85	-1.27
	Soft Teacher [7]	ICCV 2021	66.40	+0.28
	SOOD	-	**66.64**	+0.52
Oriented R-CNN [6]	Supervised	ICCV 2021	67.26	-
	SOOD	-	**68.50**	+1.24

2.3. Impact of SOOD on Stronger Detector

We adopt stronger augmentation on Oriented R-CNN [6] and evaluate the effectiveness of our SOOD, as shown in Tab. 3. Note that for evaluation on DOTA-v1.5-test, we adopt both DOTA-v1.5-train and DOTA-v1.5-val for training as in [6], using additional images from DOTA-v2.0 [1] to form unlabeled data set[1]. On DOTA-v1.5-val, SOOD improves the supervised baseline by +2.49 and reaches 71.04 mAP. Although with more labeled data for training, our SOOD can still boost the performance of the supervised baseline on DOTA-v1.5-test.

[1]We exclude the overlapped part between DOTA-v1.5 and DOTA-v2.0, using the left data of DOTA-v2.0 to form the unlabeled set.

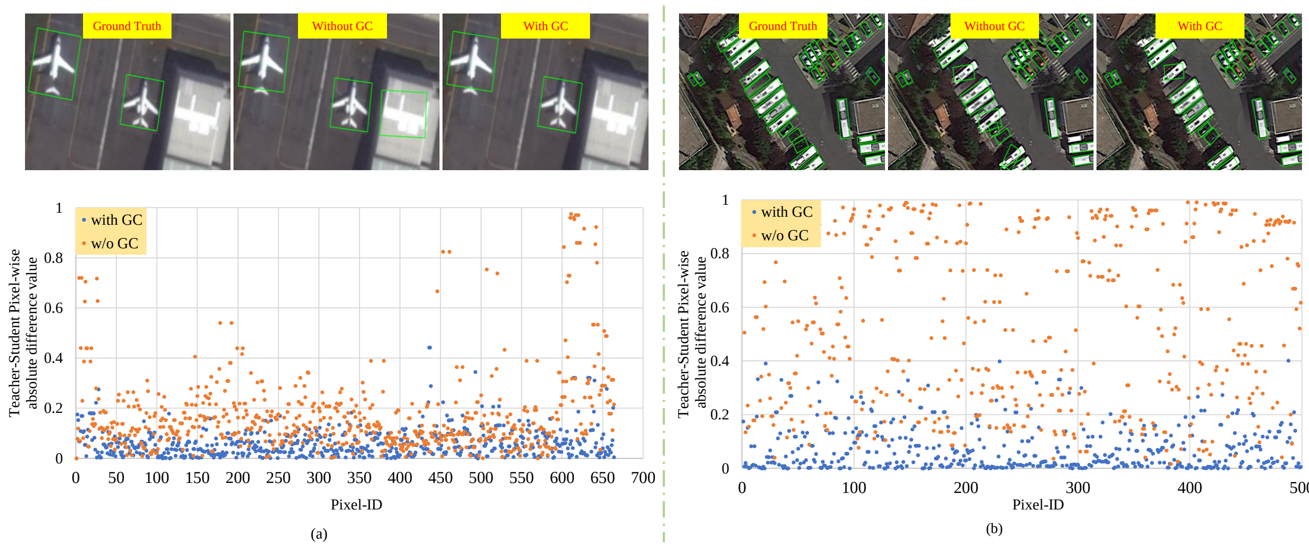


Figure 2. Visualization of the proposed GC's effect on prediction results and the distribution of absolute difference values between the teacher and student's classification scores. (a) represents sparse instances, and (b) represents dense instances. For the distribution map below, lower values indicate that the teacher and the student have more consistent distributions.

Table 3. Performance of SOOD on detector with stronger augmentation (e.g., multi-scale augmentation). The evaluation on DOTA-v1.5-test are conducted at public server[2], all methods are trained on DOTA-v1.5-train and DOTA-v1.5-val.

Detector	Method	DOTA-v1.5-val	DOTA-v1.5-test
Oriented R-CNN [6] w/ multi-scale	Supervised	68.55	75.67
	SOOD	**71.04**	**76.42**

2.4. Global Consistency Loss on Soft Teacher

We additionally adopt Global Consistency (GC) loss on Soft Teacher [7] to evaluate its generalizability. As shown in Tab. 4, GC improves Soft Teacher's performance under three settings and surpasses SOOD under the 10% setting. It indicates that GC can be easily applied to other semi-supervised paradigms.

Table 4. The performance of GC on Soft Teacher.

Setting	mAP		
	10%	20%	30%
Soft Teacher	48.46	54.89	57.83
Soft Teacher + GC (**ours**)	**49.30**	55.45	58.16
SOOD (**ours**)	48.63	**55.58**	**59.23**

3. Impact of Global Consistency Loss

We further provide qualitative visualizations to analyze the effect of the proposed Global Consistency (GC) loss. Specifically, we visualize some detection results of SOOD and SOOD without GC in Fig. 2, along with the distribution of absolute difference values between the teacher and the student's predictions. For the distribution map, lower values indicate that the teacher and the student's distributions are more consistent. Fig. 2 shows that for both sparse and dense objects, GC can improve the consistency between the teacher and the student, leading to better prediction results.

[2] https://captain-whu.github.io/DOTA/evaluation.html

References

[1] Jian Ding, Nan Xue, Gui-Song Xia, Xiang Bai, Wen Yang, Michael Ying Yang, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, et al. Object detection in aerial images: A large-scale benchmark and challenges. *IEEE transactions on pattern analysis and machine intelligence*, 44(11):7778–7796, 2021. 2

[2] Yen-Cheng Liu, Chih-Yao Ma, Zijian He, Chia-Wen Kuo, Kan Chen, Peizhao Zhang, Bichen Wu, Zsolt Kira, and Peter Vajda. Unbiased teacher for semi-supervised object detection. In *Proc. of International Conference on Learning Representations*, 2021. 2

[3] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Proc. of Advances in Neural Information Processing Systems*, 28, 2015. 2

[4] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection. In *Porc. of IEEE Intl. Conf. on Computer Vision*, pages 9627–9636, 2019. 2

[5] 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 *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, pages 3974–3983, 2018. 1

[6] Xingxing Xie, Gong Cheng, Jiabao Wang, Xiwen Yao, and Junwei Han. Oriented r-cnn for object detection. In *Porc. of IEEE Intl. Conf. on Computer Vision*, pages 3520–3529, 2021. 2, 3

[7] Mengde Xu, Zheng Zhang, Han Hu, Jianfeng Wang, Lijuan Wang, Fangyun Wei, Xiang Bai, and Zicheng Liu. End-to-end semi-supervised object detection with soft teacher. In *Porc. of IEEE Intl. Conf. on Computer Vision*, pages 3060–3069, 2021. 2, 3

[8] Hongyu Zhou, Zheng Ge, Songtao Liu, Weixin Mao, Zeming Li, Haiyan Yu, and Jian Sun. Dense teacher: Dense pseudo-labels for semi-supervised object detection. In *Proc. of European Conference on Computer Vision*, 2022. 2