

Two Teachers are Better Than One: Semi-supervised Elliptical Object Detection by Dual-Teacher Collaborative Guidance

Supplementary Material

1 MORE ON EXPERIMENTAL DEATILS

1.1 Implementation Details

In addition to the experimental details presented in the main paper, we illustrate the training loss curves of DTCG on two datasets. Specifically, the loss on the GED dataset is depicted in Fig. 1 left, where the blue, red, and green curves represent the supervised loss \mathcal{L}_{sup} , the unsupervised loss \mathcal{L}_u , and the total loss \mathcal{L}_{total} . During the early stage, as only supervised training is conducted, thus the total loss \mathcal{L}_{total} coincides with the supervised loss \mathcal{L}_{sup} , decreasing rapidly. Afterwards, when semi-supervised training \mathcal{L}_u emerges, it results in an increase in \mathcal{L}_{total} . Overall, we observe that \mathcal{L}_u rises temporarily, possibly due to instability in the initial phase. Subsequently, as training progresses, \mathcal{L}_u , \mathcal{L}_{sup} , and \mathcal{L}_{total} all eventually exhibit a steady convergence trend. Similarly, the training on the SmartPhone dataset comprises three phases, with a trend similar to that of the GED dataset, as shown in Fig. 1 right.

1.2 Evaluation Metrics

The evaluation metrics including Precision, Recall, and F-Measure are defined as follows:

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN}, \\ \text{Precision} &= \frac{TP}{TP + FP}, \\ \text{F-Measure} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \end{aligned} \quad (1)$$

where TP is the number of correctly detected ellipses, the threshold value used for determining TP is 0.8 for GED and SmartPhone datasets, FP is the number of falsely detected ellipses, and FN is the number of undetected ellipses. All these three measures target higher values for higher quality results.

1.3 More Results with Cross-dataset

In practice, it's not always guaranteed that labeled and unlabeled data originate from the same dataset. Hence, we conduct cross-dataset experiments to address this variability. In addition to the experiments detailed in the main paper, we provide supplementary results obtained from testing on another test set. The aim is to ascertain whether the model can enhance its performance by leveraging knowledge from unlabeled data sourced from a different dataset. The supplementary results are outlined in Tab. 1, including the first part of the first row and the second part of the second row. In this presentation, arrows are used to signify improvements relative to the supervised baseline, which represents the outcome of training solely with labeled data.

Specifically, in the first part of the first row, we employ the GED training set as labeled data and the SmartPhone training set as unlabeled data, resulting in a notable enhancement in our DTCG performance: 5.94% increase in Precision, 3.73% in Recall, and 3.54%

in F-Measure compared to the supervised baseline. Compared to the other two semi-supervised methods, our precision shows an improvement of at least 1.88%, recall by 0.4%, and F-Measure by 1.23%. When utilizing SmartPhone as labeled data and GED as unlabeled data, we achieve equally remarkable improvements: a 1.38% enhancement in precision, a 2.66% in recall, and a 2.07% in F-measure. This advancement notably surpasses other semi-supervised methods, demonstrating the efficacy of our approach in leveraging unlabeled data from another dataset to improve detection performance. Interestingly, we also observe a decrease in the performance of SOOD and Dense Teacher in this configuration. We speculate that this may be attributed to the simpler task of the SmartPhone dataset, potentially exacerbated by the introduction of noise from the GED dataset. This observation once again underscores the robustness inherent in our method.

1.4 Comparison with Fully-supervised methods

As reported in Table 1 of the main paper, the comparison has verified the superiority of DTCG over two semi-supervised methods. Apart from that, this experiment aims to compare DTCG with two top-performing ellipse detection methods based on fully-labeled data, including ElDet [3] and FCOS [2]. For a fair comparison, we implement these two methods using partially labeled data (like 10%, 20%, 30%), instead of a set of fully labeled data. As can be seen from the results reported in Table 2, our DTCG consistently outperforms the other two methods for various settings (10%, 20%, 30%). Specifically, on the GED dataset, we improve Precision by an average of 1.77%, Recall by an average of 4.87%, and F-Measure by an average of 4.17% compared to the second best method. Similarly, on the SmartPhone dataset, our improvement is even more significant, outperforming the methods by at least an average of 6.14% Precision, 13.66% Recall, 12.31% F-Measure, respectively. All these results reveal that leveraging additional unlabeled data is a significant solution to remedy the label scarcity and improve the representation learning.

1.5 More Qualitative Results

In addition to the qualitative results presented in the main paper, we conduct a comprehensive qualitative comparison, as illustrated in Fig. 2. Our DTCG consistently outperforms both fully supervised methods [2, 3], and semi-supervised methods [1, 4]. Specifically, DTCG demonstrates superior prediction accuracy and detection quality compared to Dense Teacher and SOOD. In the third and fourth columns of Fig. 2, both ElDet and SOOD exhibit mistakes in several targets, while FCOS and Dense Teacher produce conspicuous false-positive instances. Additionally, in the fifth column, ElDet, FCOS, and Dense Teacher fail to accurately predict angles, resulting in prediction frames that do not align well with target frames, thus potentially compromising detection accuracy. This highlights the robustness of our approach even in challenging scenarios. In a

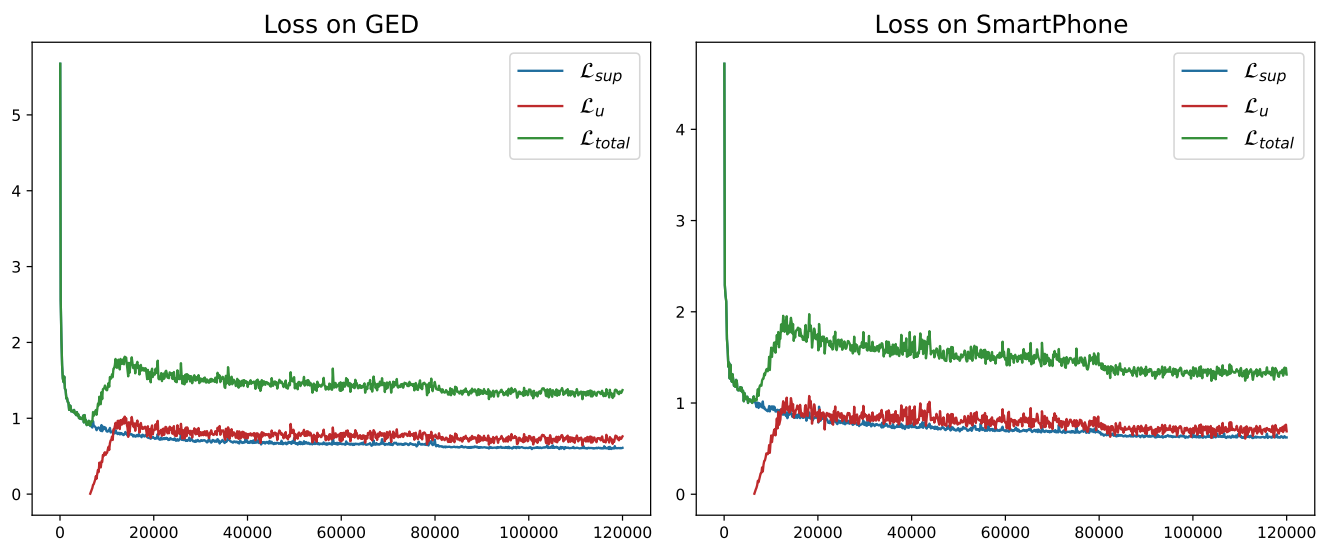


Figure 1: We show the training loss curves on the GED and SmartPhone datasets, where the x-axis represents the training iteration and the y-axis represents the loss cost.

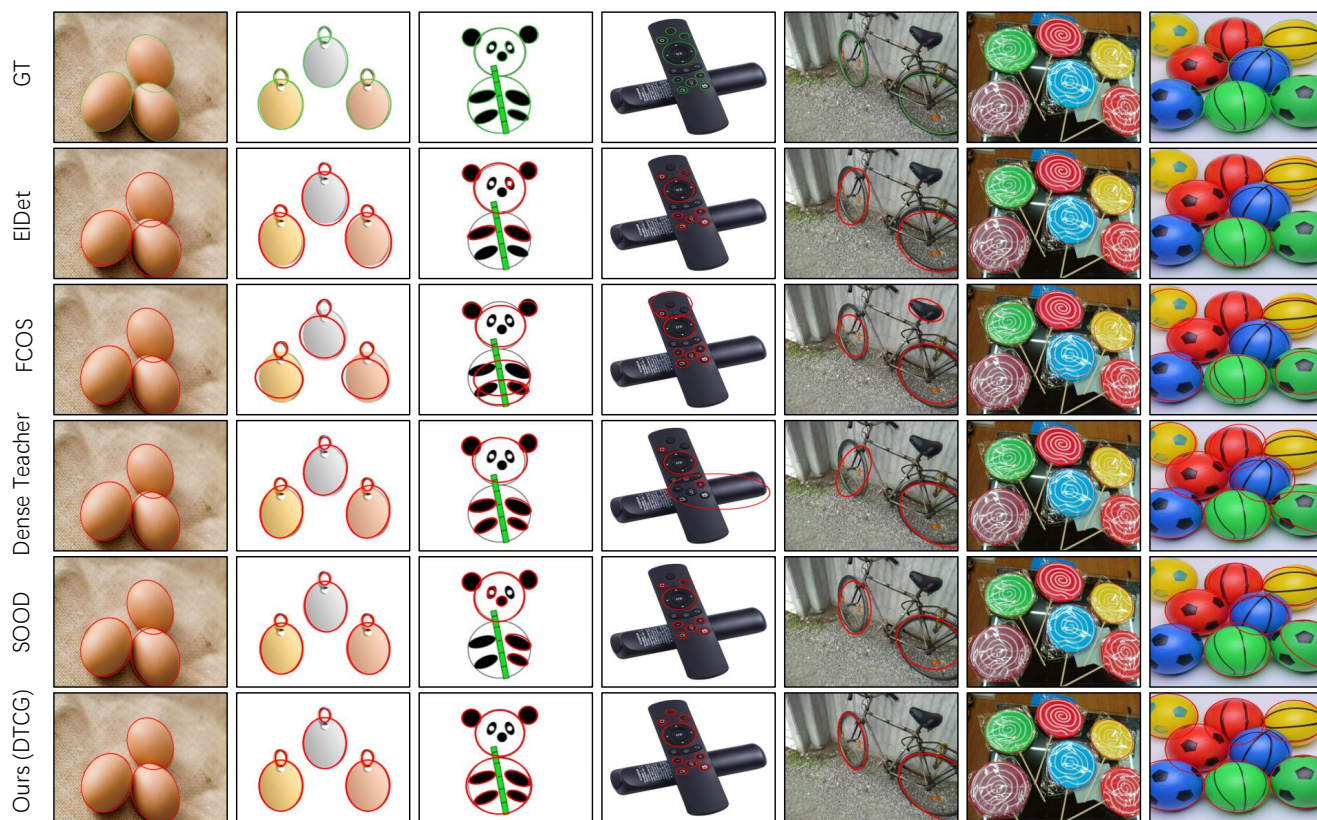


Figure 2: Qualitative comparison with other methods in the GED dataset.

Table 1: More cross-dataset evaluations using the GED and SmartPhone datasets.

Labeled	Unlabeled	Method	Test (GED)			Test (SmartPhone)		
			P	R	F-M	P	R	F-M
GED	SmartPhone	Dense Teacher[4]	77.82 $\xrightarrow{+3.38}$ 81.20	68.13 $\xrightarrow{+3.33}$ 71.46	73.72 $\xrightarrow{+2.31}$ 76.03	48.67	53.14	50.81
		SOOD [1]	77.82 $\xrightarrow{+4.06}$ 81.88	68.13 $\xrightarrow{+2.81}$ 70.94	73.72 $\xrightarrow{+2.3}$ 76.02	66.76	59.66	63.01
		DTCG (Ours)	77.82 $\xrightarrow{+5.94}$ 83.76	68.13 $\xrightarrow{+3.73}$ 71.86	73.72 $\xrightarrow{+3.54}$ 77.26	71.01	64.49	67.59
SmartPhone	GED	Dense Teacher[4]	49.05	38.36	43.05	82.47 $\xrightarrow{-8.34}$ 74.10	69.32 $\xrightarrow{-4.34}$ 64.98	75.33 $\xrightarrow{-6.09}$ 69.24
		SOOD [1]	47.50	39.12	43.10	82.47 $\xrightarrow{-3.58}$ 78.89	69.32 $\xrightarrow{-4.1}$ 65.22	75.33 $\xrightarrow{-4}$ 71.33
		DTCG (Ours)	60.81	39.66	48.01	82.47 $\xrightarrow{+1.38}$ 83.71	69.32 $\xrightarrow{+2.66}$ 71.98	75.33 $\xrightarrow{+2.07}$ 77.40

Table 2: Quantitative comparison on GED and SmartPhone datasets, under the partially labeled data settings. In addition, we show the average scores across the three settings.

Dataset	Method	10%			20%			30%			Average		
		P	R	F-M	P	R	F-M	P	R	F-M	P	R	F-M
GED	ElDet [3]	48.07	46.34	47.19	60.96	55.46	58.08	68.54	62.14	65.18	59.19	54.65	56.82
	FCOS [2]	69.66	55.35	61.84	75.38	62.35	68.25	76.21	64.90	70.10	73.75	61.87	66.73
	DTCG (Ours)	72.10	64.69	68.19	76.53	66.38	71.10	77.95	69.35	73.41	75.52	66.74	70.9
SmartPhone	ElDet [3]	67.74	45.65	54.54	87.05	58.45	69.94	75.76	60.39	67.20	76.85	54.83	63.89
	FCOS [2]	43.35	33.09	37.53	77.22	67.19	71.83	79.67	70.04	74.55	66.75	56.77	61.30
	DTCG (Ours)	78.27	67.87	72.70	80.75	67.87	73.75	89.94	75.60	82.15	82.99	70.45	76.2

notable failure case, our DTCG overlooks a single blue ball in the image of the last column, whereas the other methods miss multiple targets.

REFERENCES

[1] Wei Hua, Dingkang Liang, Jingyu Li, Xiaolong Liu, Zhikang Zou, Xiaoqing Ye, and Xiang Bai. 2023. SOOD: Towards Semi-Supervised Oriented Object Detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 15558–15567.

[2] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. 2019. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*. 9627–9636.

[3] Tianhao Wang, Changsheng Lu, Ming Shao, Xiaohui Yuan, and Siyu Xia. 2022. ElDet: An Anchor-free General Ellipse Object Detector. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*. 2580–2595.

[4] Hongyu Zhou, Zheng Ge, Songtao Liu, Weixin Mao, Zeming Li, Haiyan Yu, and Jian Sun. 2022. Dense teacher: Dense pseudo-labels for semi-supervised object detection. In *European Conference on Computer Vision*. Springer, 35–50.