Supplementary Material for *"Entropy-based Active Learning for Object Detection with Progressive Diversity Constraint"*

In this supplementary material, we elaborately analyze how different components affect the performance of the proposed diverse prototype (DivProto) strategy. In addition, we provide comparison results with the latest MDN [2] and some exploration results with a state-of-the-art detector and semi-supervised learning. To make a more comprehensive evaluation, we present additional experimental results on MS COCO as complements to Fig. 3 of the main paper.

A. Ablation Study of DivProto

As depicted in Section 4.3, DivProto consists of intra-class redundancy rejection and inter-class balanced selection. We separately evaluate the effects of these two parts, with the Basic Entropy as the baseline for comparison. As summarized in Table A, both intra-class rejection and inter-class balanced selection improve the performance of the baseline under various annotation percentages. Their combination further promotes the AP with 25% and 35% annotated percentages and remains highly competitive in the other cases.

Method	Annotated Percentage				
	20%	25%	30%	35%	40%
Basic Entropy	27.57	29.38	30.61	31.47	32.36
Intra-class	27.57	29.52	30.32	31.15	**32.57**
Inter-class	27.57	29.59	**30.70**	31.83	32.37
Both	27.57	**29.73**	30.64	**31.86**	32.53

Table A. AP (%) on MS COCO by using intra-class redundancy rejection, inter-class balanced selection and their combination, compared to the Basic Entropy baseline. All the methods are based on Faster R-CNN with the ResNet-50 backbone. The best result for each method is highlighted in bold.

B. Comparison with the latest work MDN

MDN [2] delivers gains in two ways: an acquisition method based on uncertainty disentanglement and an improved SSD detector based on GMM. In contrast, our method mainly focuses on acquisition, and ENMS and DivProto are proposed to handle redundant uncertainty estimation and insufficient cross-image diversity, both of which are not considered in [2]. To eliminate the effect of the detector, we apply our acquisition method to the improved SSD on VOC07+12 with the same setting as [2]. As in Table B, our method outperforms [2], showing its advantage in acquisition.

Acquisition Method	Detector	2k	3k	4k
MDN [2]	Improved SSD [2]	61.30	66.57	68.49
Ours	Improved SSD [2]	63.35	67.56	70.33

Table B. mAP (%) of different acquisition methods on VOC07+12.

C. Results on SOTA Detectors

Our method is designed for active acquisition independent of detectors, thus theoretically being effective for different detectors. We additionally evaluate our method by using YOLOv5[1] in Table C, showing its effectiveness with SOTA detectors.

Method	2%	4%	6%	8%	10%
Random	9.96	15.93	19.96	23.37	25.06
Ours	9.96	16.68	21.03	24.06	26.39

Table C. AP (%) using the YOLOv5 detector on MS COCO.

D. Combining with Semi-supervised Learning

In our opinion, semi-supervised learning is similar to active learning in the goal of pursuing better performance with less annotated data but in different learning paradigms. As in Table D, we achieve better results by combining our method with a reputed semi-supervised one, UBT [3], where they complement each other.

Method	2%	4%	6%	8%	10%
Ours	6.70	15.44	18.82	20.83	22.26
UBT	24.30	27.01	28.45	29.41	31.50
Ours+UBT	24.30	28.55	30.28	31.70	32.24

Table D. AP (%) by combining UBT on MS COCO.

[1] https://github.com/ultralytics/yolov5

E. Comprehensive Results on MS COCO

In Fig. 3 of the main paper, we report the Average Precision (AP) over IoU thresholds from 0.5 to 0.95 on MS COCO, by using our method as well as the counterpart methods: Core-set [4], CDAL [1], Learn Loss [5], and MIAL [6]. In Table E, we provide more comparison results under different metrics, including AP_{50}, AP_{75}, AP_S, AP_M, and AP_L. Here, AP_{50}, AP_{75} are AP at the 0.5 and 0.75 IoU thresholds, respectively. AP_S, AP_M, and AP_L indicate AP for small, medium, and large objects, respectively.

As shown in Table E, our method remarkably outperforms the counterparts in most cases. It is worth noting that Core-set [4] performs better than ours at detecting large objects, since it adopts spatial pooling to merge instance-level features to a holistic image-level representation, based on which the significance of large objects is strengthened. By contrast, our method employs a more balanced way to integrate instance-level features, thus achieving a higher averaged AP for all scales at the cost of slightly lower AP for large objects.

References

[1] Sharat Agarwal, Himanshu Arora, Saket Anand, and Chetan Arora. Contextual diversity for active learning. In *European Conference on Computer Vision*, pages 137–153, 2020. 2

[2] J. Choi and et al. Active learning for deep object detection via probabilistic modeling. In *IEEE International Conference on Computer Vision*, 2021. 1

[3] Y. Liu and et al. Unbiased teacher for semi-supervised object detection. In *International Conference on Learning Representations*, 2021. 1

[4] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. In *International Conference on Learning Representations*, 2018. 2

[5] Donggeun Yoo and In So Kweon. Learning loss for active learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 93–102, 2019. 2

[6] Tianning Yuan, Fang Wan, Mengying Fu, Jianzhuang Liu, Songcen Xu, Xiangyang Ji, and Qixiang Ye. Multiple instance active learning for object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 2

Images	Method	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
20%	Random	27.57	48.81	28.09	14.53	29.87	36.12
25%	Random	28.97	50.43	29.84	15.39	31.25	37.84
	Core-set	28.75	50.00	29.56	14.59	30.98	**38.79**
	CDAL	29.11	48.60	27.86	14.29	29.69	35.82
	Learn Loss	29.42	51.08	30.38	16.35	31.96	37.81
	MIAL	29.39	51.21	30.36	15.96	31.87	38.68
	Entropy	29.38	51.06	30.31	16.18	31.63	37.89
	ENMS	29.76	51.64	30.78	**16.72**	32.12	38.03
	DivProto	29.73	51.45	**30.84**	16.62	32.32	38.08
	Ours	**29.78**	**51.70**	30.81	16.52	**32.32**	38.00
30%	Random	30.07	51.61	31.13	16.05	32.42	39.42
	Core-set	29.90	51.37	31.12	15.60	32.19	**40.65**
	CDAL	30.01	51.57	31.27	16.28	32.72	39.21
	Learn Loss	30.55	52.53	31.90	**17.68**	33.10	38.52
	MIAL	30.47	52.49	31.45	16.85	33.44	39.31
	Entropy	30.61	52.66	31.96	17.49	33.13	38.97
	ENMS	30.82	52.89	**32.07**	17.40	**33.48**	38.88
	DivProto	30.64	52.64	31.77	17.21	33.37	38.92
	Ours	**30.90**	**53.08**	32.01	17.56	33.44	39.10
35%	Random	30.99	52.70	32.47	16.88	33.52	40.35
	Core-set	30.69	52.25	31.97	15.96	33.17	**41.76**
	CDAL	31.17	53.07	32.55	17.22	33.84	40.56
	Learn Loss	31.19	53.19	32.67	17.48	34.04	39.23
	MIAL	31.75	53.89	33.42	17.51	**34.60**	41.09
	Entropy	31.47	53.59	32.95	18.01	34.16	39.76
	ENMS	31.79	54.05	33.29	18.11	34.50	40.14
	DivProto	31.86	54.08	33.41	**18.24**	34.58	40.56
	Ours	**31.99**	**54.18**	**33.51**	18.09	34.39	40.54
40%	Random	31.62	53.29	33.18	17.14	34.19	41.26
	Core-set	31.31	52.89	32.80	16.19	33.81	**42.85**
	CDAL	31.75	53.57	33.38	17.75	34.53	41.23
	Learn Loss	32.33	54.61	33.92	18.72	35.37	40.76
	MIAL	32.27	54.69	34.04	17.72	35.27	41.86
	Entropy	32.36	54.69	33.97	18.67	35.25	40.63
	ENMS	32.56	54.84	34.25	18.53	35.42	40.96
	DivProto	32.53	54.77	34.24	18.57	35.34	41.44
	Ours	**32.87**	**55.15**	**34.58**	**18.99**	**35.57**	41.44

Table E. AP by using various active learning based methods on MS COCO. All the results are based on Faster R-CNN with the ResNet-50 backbone. AP_{50} and AP_{75} refer to AP at the 0.5 and 0.75 IoU thresholds. AP_S, AP_M, and AP_L indicate AP for objects with small, medium, and large sizes, respectively.