Improving Single Domain-Generalized Object Detection: A Focus on Diversification and Alignment

Supplementary Material

1. Additional Results

Sensitivity to hyperparameters: We analyze the sensitivity of α and β (from Eq.(6) in the main paper) in Tables 1 and 2, respectively. We observe that overall the performance of our method is robust to the choice of hyperparameters. It outperforms the baseline on all different settings of hyperparameters.

α	Clipart	Watercolor	Comic
0.0	35.0	53.8	28.7
0.1	36.1	53.6	27.1
0.3	37.9	55.1	30.2
0.5	38.0	55.7	29.7
0.7	37.1	55.6	31.5
1.0	**38.9**	**57.4**	**33.2**
1.2	38.6	56.8	33.0

Table 1. Sensitivity analysis of hyperparameter α after setting β to a fixed value of 1.0.

β	Clipart	Watercolor	Comic
0.0	36.2	53.9	28.7
0.1	36.1	54.8	30.2
0.3	37.8	55.7	30.8
0.5	**39.4**	54.7	32.0
0.7	37.3	56.8	32.2
1.0	38.9	**57.4**	**33.2**
1.2	37.7	56.3	31.7

Table 2. Sensitivity analysis of hyperparameter β after setting α to 1.0.

Results with another diversification approach: A recent work dubbed as Normalization Perturbation (NP) [2], which targets single-DGOD, proposed to diversify the single domain by perturbing the low-level channel statistics. We evaluated this method on Real to Artistic benchmark and report the result in Table 3. To further investigate the efficacy of our alignment losses, namely $\mathcal{L}_{\text{cal}} + \mathcal{L}_{\text{ral}}$ from sec 3.2.2 in the main paper, we replace our diversification technique (sec 3.2.1 main paper) with the one proposed in [2]. Table 3 shows that our alignment losses are effective even with a different diversification approach, however, the results are still inferior to our proposed method.

Method	VOC	Clipart	Watercolor	Comic
NP	79.2	35.4	53.3	28.9
NP + $\mathcal{L}_{\text{cal}} + \mathcal{L}_{\text{ral}}$	77.9	37.3	56.0	32.2
Ours	**80.1**	**38.9**	**57.4**	**33.2**

Table 3. Results after replacing our diversification technique (sec 3.2.1 main paper) with the one proposed in NP [2]. The model is trained on Pascal VOC and tested on Clipart1k, Watercolor2k and Comic2k.

Results on Medical Dataset: We also evaluate our method under the domain shift in a medical imaging scenario (see Tables 4,5). To this end, we show results on recently proposed Malaria detection M5 benchmark[4]. This benchmark includes data from two domains i.e. images of blood-smear from different malaria patients taken by a high cost microscope (HCM) and a low cost microscope i.e (LCM). The HCM serves as the source domain and LCM is used as the target domain. Table 4 shows that our proposed method is capable of generalizing to an unseen medical imaging domain (LCM), and it outperforms several other domain-adaptive detectors. Also, Table 5 reveals that, compared to baseline (Faster R-CNN) and only diversification, our method improves the calibration of out-of-domain detections.

Method	HCM	LCM
Xu et al.[5]	-	15.5
Saito et al.[3]	-	24.8
Chen et al.[1]	-	17.6
Sultani et al.[4]	66.8	**37.5**
Faster R-CNN	71.4	15.1
Diversification	**74.7**	25.0
Our Method	70.3	**35.9**

Table 4. Performance comparison (mAP %) under domain shift in medical imaging modality. Model is trained on HCM and tested on LCM (out-domain).

Method	LCM
Faster R-CNN	8.4
Diversification	8.0
Ours	**5.5**

Table 5. Calibration performance using D-ECE metric (%) under domain-shift in medical imaging modality.

Additional results with FCOS: We also report results of our method with FCOS baseline on Urban scene detection benchmark (see Table 6). We note that, our method is also effective with FCOS baseline on the challenging multi-weather domain shifts.

Method	DS	NC	DR	NR	DF
FCOS	40.8	29.7	22.1	12.6	25.3
Diversification	48.3	35.6	32.6	18.7	31.3
Ours	**53.7**	**37.9**	**37.4**	**21.2**	**34.9**

Table 6. Results (%) of our method with FCOS baseline on multi-weather scenario where model is trained on Daytime Sunny (DS) and tested on Night-Clear (NC), Night-Rainy (NR), Dusk-Rainy (DR) and Daytime-Foggy (DF).

Diversification with constant amplitude: We omit the constant amplitude augmentation method from our final pool of visual corruptions for achieving diversification as it destroys the instance-level information which is crucial for object detection. Fig. 1 shows some sample images when constant amplitude is applied. We see that the objects information is mostly destroyed.

Additional qualitative results: We sample some results from urban scene detection benchmark in Fig. 2. Our alignment losses allows detecting the object that were missed by baselines i.e Faster R-CNN and our diversification techniques as shown in columns (c, f). Diversification result in more false positive detection as shown in column (e). For instance, in Night Rainy, Daytime Foggy and Night Clear, the background is detected as car and bus, and in dusk rainy, a bus is misclassfied as car.

Class-wise performance on Clipart1k: Table 7 shows the class-wise results when trained on Pascal VOC and evaluated on Clipart1k.

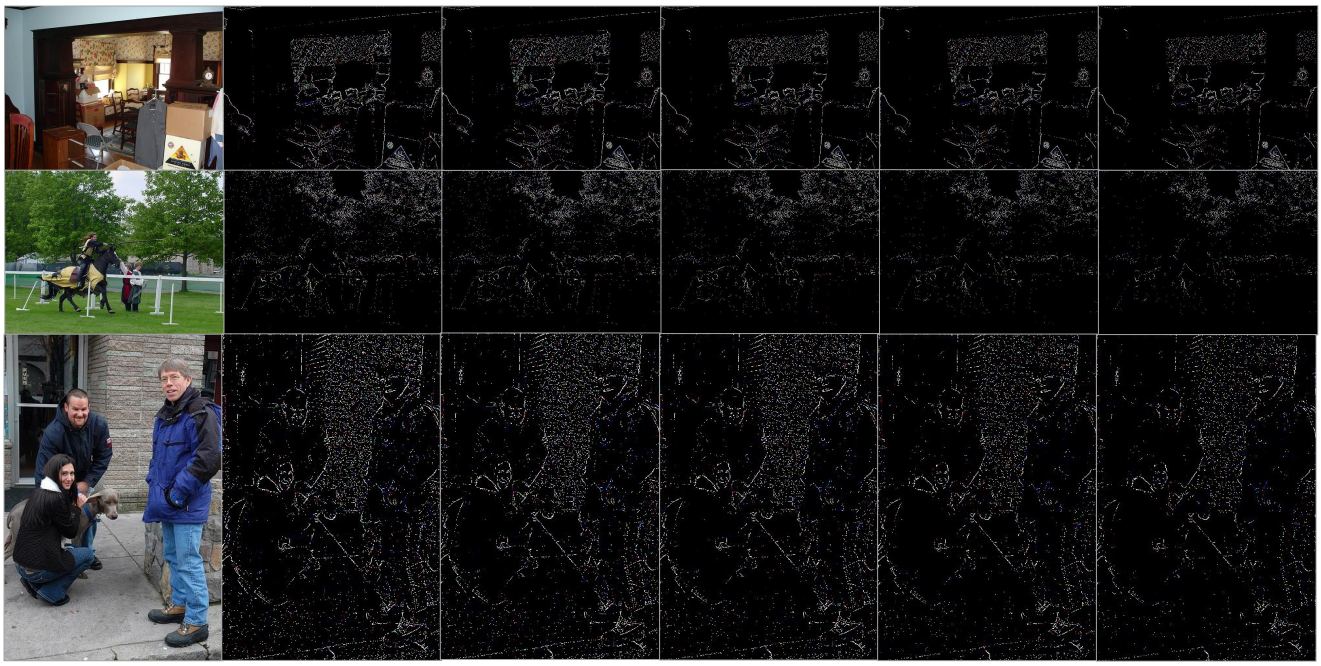

Figure 1. Diversification results after applying constant amplitude visual corruption. We see that the objects information is mostly destroyed. Here the columns show the increasing severity levels.

Figure 2. Sampled results from urban scene detection benchmark. Our method accurately detects objects e.g car, truck, person, bus. For better viewing, higher zoom is recommended.

Method	place	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor.	person	plant	sheep	sofa	train	tv	mAP
Faster R-CNN	19.9	51.6	17.0	21.9	27.2	49.6	25.5	9.1	35.1	9.1	25.4	3.0	29.2	48.9	30.1	40.3	9.1	6.7	35.2	21.0	25.7
Diversification (div.)	29.3	50.9	23.4	**35.3**	45.3	49.8	**33.4**	**10.6**	43.3	22.3	31.6	4.5	32.9	51.9	40.2	51.1	18.2	29.6	42.3	28.5	33.7
div. + $\mathcal{L}_{\text{cal}}$	30.4	61.7	23.8	28.4	41.6	51.7	32.6	7.6	**46.9**	31.4	**36.8**	8.3	41.7	57.2	48.4	48.5	18.2	38.2	40.1	30.4	36.2
div. + $\mathcal{L}_{\text{ral}}$	34.0	56.2	**24.4**	28.9	41.2	50.4	32.6	4.5	45.9	29.5	30.2	12.8	39.4	**59.5**	46.4	50.1	**18.2**	26.2	**43.4**	26.2	35.0
div. + $\mathcal{L}_{\text{cal}} + \mathcal{L}_{\text{ral}}$ (Ours)	**34.4**	**64.4**	22.7	27.0	**45.6**	**59.2**	32.9	7.0	46.8	**55.8**	28.9	**14.5**	**44.4**	58.0	**55.2**	**52.1**	14.8	**38.4**	42.5	**33.9**	**38.9**

Table 7. Class-wise AP(%) comparison of baseline and proposed method on Pascal VOC to Clipart1k scenario.

References

[1] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3339–3348, 2018. 1

[2] Qi Fan, Mattia Segu, Yu-Wing Tai, Fisher Yu, Chi-Keung Tang, Bernt Schiele, and Dengxin Dai. Towards robust object detection invariant to real-world domain shifts. In *The Eleventh International Conference on Learning Representations*, 2023. 1

[3] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Strong-weak distribution alignment for adaptive object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6956–6965, 2019. 1

[4] Waqas Sultani, Wajahat Nawaz, Syed Javed, Muhammad Sohail Danish, Asma Saadia, and Mohsen Ali. Towards low-cost and efficient malaria detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20687–20696, 2022. 1

[5] Minghao Xu, Hang Wang, Bingbing Ni, Qi Tian, and Wenjun Zhang. Cross-domain detection via graph-induced prototype alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12355–12364, 2020. 1