Robust Object Detection via Instance-Level Temporal Cycle Confusion
Supplementary Materials

In this supplementary material, we provide the detailed results and analysis of the out-of-domain evaluation benchmarks. In Section 1, we also show the per-category results of the benchmark results to complement the results in the Table 2 and Table 3 of the main paper and other related evaluations on the Waymo open dataset. In Section 2, we provide more visualizations of the prediction results.

1. Out-of-Domain Evaluation Results

In this section, we provide extensive evaluation results of our approach on various datasets and evaluation settings. We find our approach consistently improve over the baseline approaches and the evaluation results on a range of datasets and evaluation settings verify the generality of our method.

Per-category results on BDD100K and Waymo. In this section, we provide per-category evaluation results to complement the Table 2 and 3 in the main paper. We also indicate the number of instances in each category to give the readers an idea of the evaluation data distribution.

In Table 1 and Table 2, we provide the per-category evaluation results for BDD100K Daytime → Night and BDD100K Night → Daytime, which is originally shown in Table 2 of the main paper. Our CycConf task is able to outperform all other tasks across both settings. On BDD100K Daytime → Night, CycConf can achieve at least 1 point improvement in both AP50 and AP75 over other methods. We can observe improvements in several rare categories, including bus, truck, bicycle, and motorcycle. On BDD100K Night → Daytime, CycConf can achieve significant improvements on car and motorcycle categories.

In Table 3 and Table 4, we provide the per-category results for Waymo Front Left → BDD100K Night and Waymo Front Right → BDD100K Night to complement the results in Table 3 of the main paper. Across both settings, CycConf outperforms all other methods by a large margin. On Waymo Front Left → BDD100K Night, CycConf can achieve at least 1.5 points improvement in both AP50 and AP75 over other methods, improving the performance on vehicles and pedestrians. On Waymo Front Right → BDD100K Night, CycConf improves the performance on both vehicles and cyclists, while achieving competitive performance with FRCN + Rot on pedestrians.

Additional cross-camera evaluation on Waymo. In Table 5 and Table 6, we provide results for additional settings on the Waymo dataset, Waymo Front → Front Left and Waymo Front → Side Left. In these settings, the domain gap is due to the change in camera angles. We observe significant improvements for CycConf on both settings. On Waymo Front → Front Left, CycConf can achieve around 2 points improvement in AP50 and AP75 and around 4 points improvement in AP for cyclists, while achieving competitive performance on AP for vehicles and pedestrians. The other self-supervised methods can not obtain improvements. On Waymo Front → Side Left, CycConf can achieve around 1 point improvement in AP75 and in AP for cyclists.

2. Visualizations of Prediction Results

We visualize predictions of CycConf trained on BDD100K Daytime on frames from several video sequences of BDD100K Night in Figure 1. Although our model does not observe nighttime images during training, it is still able to successfully identity a majority of the ground truth labels, especially in the densely populated areas.

In Figure 2, we show the visualization of UDA experiment on Cityscape dataset. From the left to right, we provide the prediction results of the baseline model, the detector trained with rotation and the ground truths. We can observe a more robust prediction of our model under severe distribution shifts.

Table 1: BDD100K Daytime → Night (complement to the result of Table 2 in the main paper).

Model	AP	AP50	AP75	person	rider	car	bus	truck	bicycle	motorcycle	train
# Instances				12606	737	107531	1760	4033	846	130	63
FRCN	17.84	31.34	17.68	30.62	13.19	41.39	14.36	23.38	11.46	8.38	0.00
+Rot	18.58	32.95	18.15	**30.76**	**14.39**	41.38	14.17	23.07	11.62	13.24	0.00
+Jigsaw	17.47	31.22	16.81	29.86	13.05	41.24	14.07	21.91	11.23	8.38	0.00
+Cycle Consist.	18.35	32.44	18.07	30.19	12.61	**42.57**	15.49	22.82	11.03	12.12	0.00
+CycConf	**19.09**	**33.58**	**19.14**	30.68	13.73	41.73	**16.71**	**24.35**	**12.00**	**13.53**	0.00

Table 2: BDD100K Night → Daytime (complement to the results of Table 2 in the main paper).

Model	AP	AP50	AP75	person	rider	car	bus	truck	bicycle	motorcycle	train
# Instances				41886	1695	200372	6110	21274	3047	770	245
FRCN	19.14	33.04	19.16	29.63	12.90	46.55	22.12	16.82	14.03	11.04	0.00
+Rot	19.07	33.25	18.83	29.61	13.92	46.70	22.67	16.29	**14.10**	9.30	0.00
+Jigsaw	19.22	33.87	18.72	30.03	13.68	47.01	21.68	16.49	13.94	10.97	0.00
+Cycle Consist.	18.89	33.50	18.31	**30.12**	13.21	47.13	22.05	17.43	13.24	7.96	0.00
+CycConf	**19.57**	**34.34**	**19.26**	29.83	**13.95**	**47.80**	**23.54**	**17.10**	11.58	**12.80**	0.00

Table 3: Waymo Front Left → BDD100K Night (complement to the results in Table 3 of the main paper).

Model	AP	AP50	AP75	vehicle	pedestrian	cyclist
# Instances				123749	12884	737
FRCN	10.07	19.62	9.05	19.41	10.12	0.69
+Rot	11.34	23.12	9.65	21.33	12.02	0.68
+Jigsaw	9.86	19.93	8.40	20.40	8.78	0.41
+Cycle Consist.	11.55	23.44	10.00	22.34	11.27	1.04
+CycConf	**12.27**	**26.01**	**10.24**	**22.91**	**12.70**	**1.19**

Table 4: Waymo Front Right → BDD100K Night (complement to the results in Table 3 of the main paper).

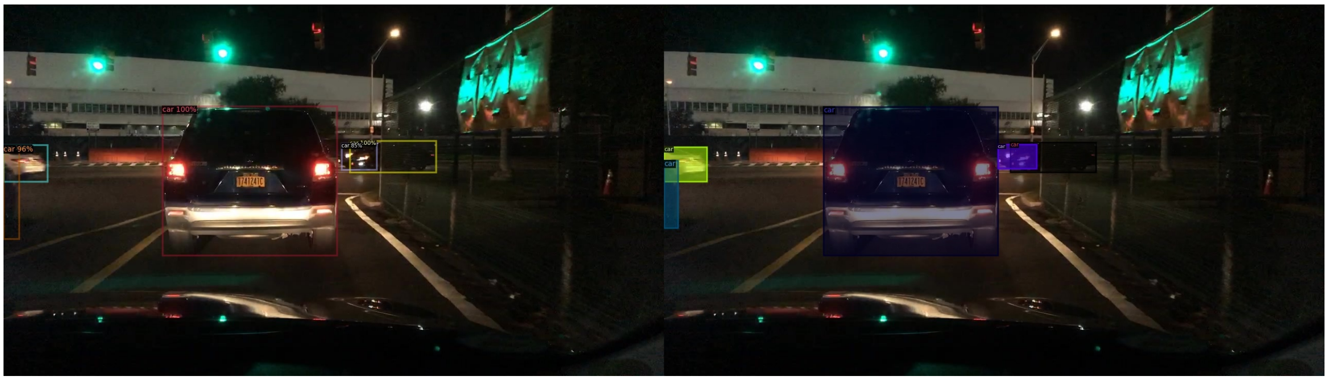
Model	AP	AP50	AP75	vehicle	pedestrian	cyclist
# Instances				123749	12884	737
FRCN	8.65	17.26	7.49	17.89	7.64	0.42
+Rot	9.25	18.48	8.08	18.22	**9.28**	0.26
+Jigsaw	8.34	16.58	7.26	16.64	8.25	0.13
+Cycle Consist.	9.11	17.92	7.98	18.87	7.80	0.65
+CycConf	**9.99**	**20.58**	**8.30**	**20.09**	9.15	**0.73**

Table 5: Waymo Front → Front Left.

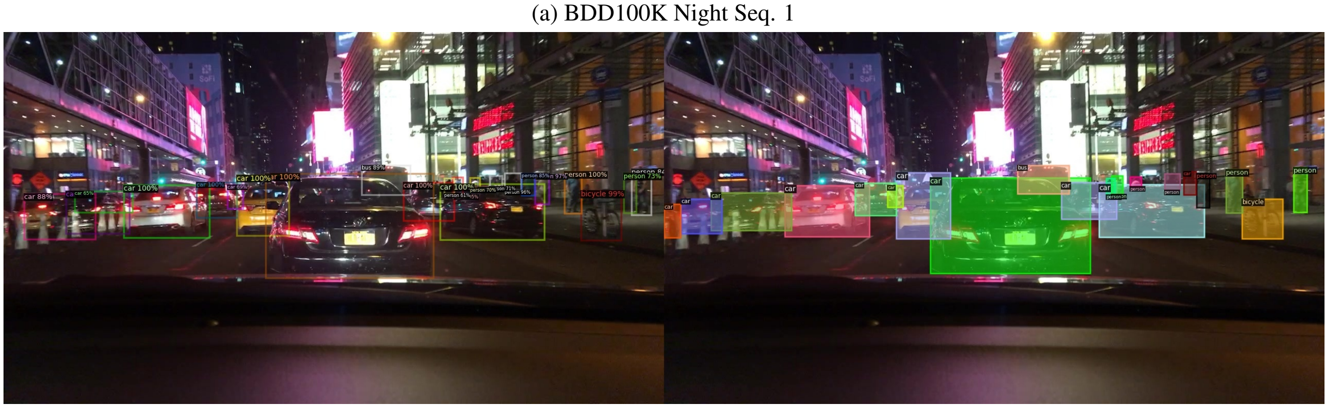
Model	AP	AP50	AP75	vehicle	pedestrian	cyclist
# Instances				297909	87221	1518
FRCN	36.05	57.73	38.27	**42.08**	36.99	29.08
+Rot	35.96	57.82	38.33	41.86	36.87	29.16
+Jigsaw	35.89	57.54	38.21	41.80	**37.05**	28.81
+Cycle Consist.	35.44	56.75	37.79	41.90	36.89	27.51
+CycConf	**37.35**	**59.78**	**40.25**	41.98	36.91	**33.15**

Table 6: Waymo Front $\rightarrow$ Side Left.

Model	AP	AP50	AP75	vehicle	pedestrian	cyclist
# Instances				283889	52938	1001
FRCN	31.92	53.95	33.31	36.69	29.57	29.48
+Rot	32.66	54.29	34.33	**36.75**	29.49	31.75
+Jigsaw	32.18	53.38	33.50	**36.75**	**29.78**	30.02
+Cycle Consist.	31.81	53.56	33.38	36.72	29.40	29.32
+CycConf	**32.89**	**54.56**	**35.27**	36.42	29.51	**32.74**



(a) BDD100K Night Seq. 1



(b) BDD100K Night Seq. 2

(c) BDD100K Night Seq. 3

Figure 1: Visualization of the predictions of Faster R-CNN w/ CycConf from BDD100K Daytime to Night. The left side is the model prediction and the right side is the ground truth labels.

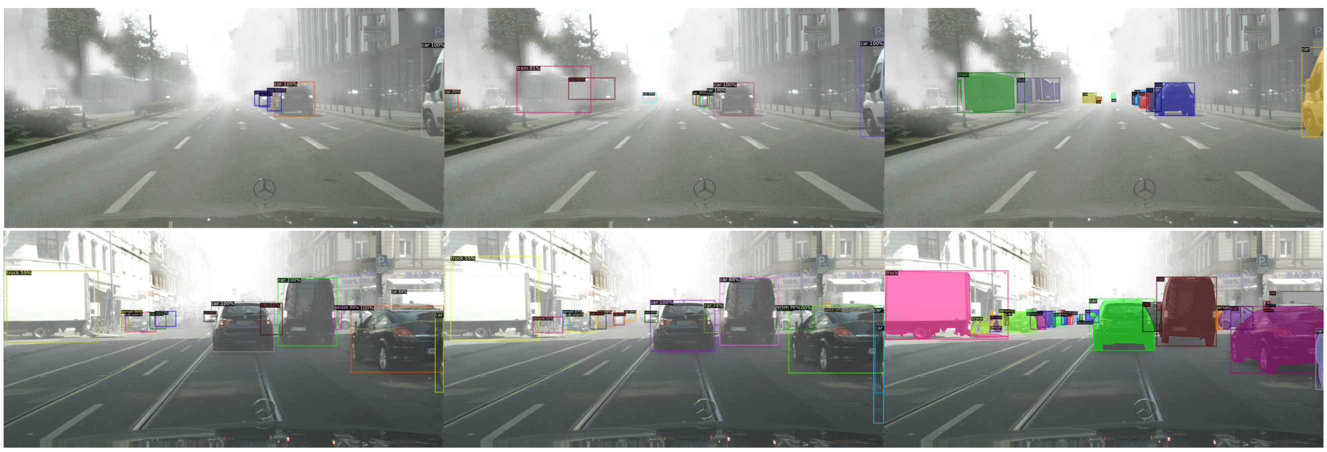


Figure 2: Visualization of UDA experiment on Cityscape dataset. From left to right are the results of baseline, w/rotation and ground-truth. We can observe the model trained with self-supervised task can better find the small or distant objects in the foggy weather, which may be ignored by the baseline methods.