Supplementary Material for "MeGA-CDA: Memory Guided Attention for Category-Aware Unsupervised Domain Adaptive Object Detection"

1. Network architecture

In Table 2, we show the architecture of the global ($\mathcal{D}_{gda}$) and category-wise ($\mathcal{D}^k_{cda}$) discriminators. The architecture of the networks Θ_t and Θ^k_t are similar as shown in Table 3.

2. Alternative baselines

Table 1 provides additional results corresponding to the ablation analysis of multi-discriminators at the instance and image level. We have employed MADA [1] algorithm in object detection to study the performance of category-wise instance-level domain adaptation. MADA [1] is implemented by category-wise domain adaptation loss at instance level multiplied with that predicted class's softmax probability for both source and target domains. Furthermore, at the image-level, we analyze the global alignment between k-global discriminator and category-wise discriminator. By studying these two methods, we can argue category-wise adaptation is crucial for category-wise alignment for both source and target domain. In the k-global discriminator experiment, k discriminators are applied on the image level.

From Table 1 we can infer that by doing category-wise domain alignment (GDA+CDA+MA) at the image level, mAP is increased by 12% when compared to category-wise domain alignment at the instance level (MADA). This shows that category-specific alignment approaches used for adapting classification models do not directly translate well for detection models. Furthermore, adding k multi-discriminator at image level perform 2% lower than category-wise domain alignment at the image level. This shows that blindly adding more discriminators do not have the ability to route specific information through each discriminator to boost overall feature alignment. Adding category-wise k discriminators as proposed in the paper would achieve that by routing only category-specific information to respective discriminators.

Table 1. Ablation analysis of multi-discriminators at instance and image level for Cityscapes $\rightarrow$ Foggy-Cityscapes dataset. k- Number of classes

Method	person	rider	car	truck	bus	train	mcycle	bicycle	mAP
Source Only	25.8	33.7	35.2	13.0	28.2	9.1	18.7	31.4	24.4
MADA [1]	28.4	34.7	35.6	15.8	32.1	10.4	20.6	32.3	26.3
k Multi-disc	35.5	44.9	51.1	26.5	40.8	24.6	36.3	37.9	37.2
GDA+CDA+MA	36.3	45.5	51.7	25.1	44.8	35.8	32.7	37.9	38.7
Oracle	37.2	48.2	52.7	35.2	52.2	48.5	35.3	38.8	43.5

Table 2. Global and Category-wise domain discriminator

Gradient Reversal Layer
Conv, 1×1, 64, stride 1, ReLU
Conv, 3×3, 64, stride 1, ReLU
Conv, 3×3, 64, stride 1, ReLU
Conv, 3×3, 3, stride 1

Table 3. Learned Similarity

Conv, 3×3, 512, stride 1, ReLU
Conv, 3×3, 256, stride 1, ReLU
Conv, 3×3, 128, stride 1, ReLU
Conv, 1×1, 64, stride 1

3. Qualitative results

References

[1] Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-adversarial domain adaptation. *arXiv preprint arXiv:1809.02176*, 2018. 1

DAFaster MeGA-CDA

Figure 1. More detection visualization for Foggy→Cityscapes. Left: DAFaster RCNN, Right: Proposed method. The bounding boxes are colored based on the detector confidence using the color map as shown. From the above visualization, we can infer that our model efficiently tackles classes' negative transfer and constructs high confidence prediction boxes. We show detections with scores higher than 0.5.

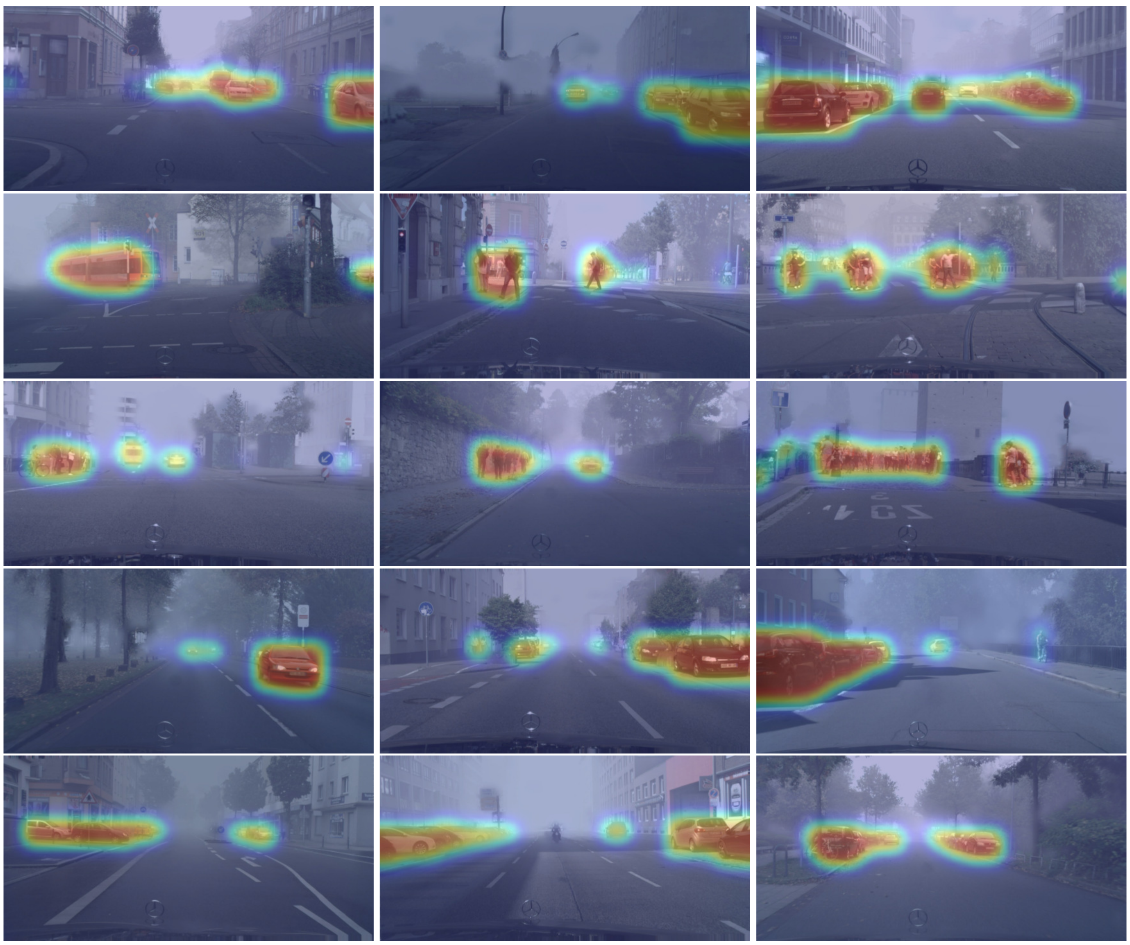

Figure 2. More visualizations of memory guided attention maps on target domains