Supplementary materials
ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation

Tuan-Hung Vu[1] Himalaya Jain[1] Maxime Bucher[1] Matthieu Cord[1,2] Patrick Pérez[1]

[1]valeo.ai, Paris, France [2]Sorbonne University, Paris, France

In this document, we provide supplementary materials for our CVPR 2019 paper. Section 1 details our entropy-based UDA models for object detection. In Section 2, we present more qualitative results for semantic segmentation and detection.

1. Entropy-based UDA for object detection

Object detection framework. We use the Single Shot MultiBox Detector (SSD-300) [1] with the VGG-16 base CNN [2] as the detection backbone in our experiments. Given an input image, SSD-300 produces dense predictions from M feature maps at different resolutions. In detail, every location on each feature map m corresponds to a set of K_m anchor boxes with predefined aspect ratios and scales. The detection pipeline ends with a non-maximum suppression (NMS) step to post-process the predictions. Readers are referred to [1] for more details about the architecture and the training procedure. We denote the "soft-detection map" of the SSD model at feature map m of dimension $H_m \times W_m$ as $\boldsymbol{P}_{\boldsymbol{x}}^m \in [0,1]^{H_m \times W_m \times K_m \times C}$.

Direct entropy minimization. Considering a target input image $\boldsymbol{x}_t$, the entropy map produced at a feature map m, $\boldsymbol{E}_{\boldsymbol{x}_t}^m \in [0,1]^{H_m \times W_m \times K_m}$, is composed of the independent box-level entropies normalized to $[0,1]$:

$$\boldsymbol{E}_{\boldsymbol{x}_t}^{m(h,w,k)} = \frac{-1}{\log(C)} \sum_{c=1}^{C} \boldsymbol{P}_{\boldsymbol{x}_t}^{m(h,w,k,c)} \log \boldsymbol{P}_{\boldsymbol{x}_t}^{m(h,w,k,c)}. \tag{1}$$

The entropy loss $\mathcal{L}_{ent}$ is defined as the sum of normalized box entropies over all anchor boxes and all feature resolutions:

$$\mathcal{L}_{ent}(\boldsymbol{x}_t) = \sum_{m} \sum_{h,w,k} \boldsymbol{E}_{\boldsymbol{x}_t}^{m(h,w,k)}. \tag{2}$$

Similar to the semantic segmentation task, we jointly optimize the supervised object detection losses on source and the unsupervised entropy loss $\mathcal{L}_{ent}$ on target samples.

Entropy minimization with adversarial learning. We apply the adversarial framework proposed in Section 3.2. To this end, we first transform the soft-detection maps $P_{\boldsymbol{x}}^m$ to the weighted self-information maps $I_{\boldsymbol{x}}^m$. We then zero-pad the $I_{\boldsymbol{x}}^m$ maps at lower resolutions to match the size of the largest one. Finally, we stack all zero-padded $I_{\boldsymbol{x}}^m$ to produce $I_{\boldsymbol{x}}$, which serves as the input to the discriminator.

2. Qualitative examples

Figure 1 and Figure 2 illustrate qualitative results for semantic segmentation and object detection.

References

[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg. SSD: single shot multibox detector. In *ECCV*, 2016. 1

[2] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015. 1

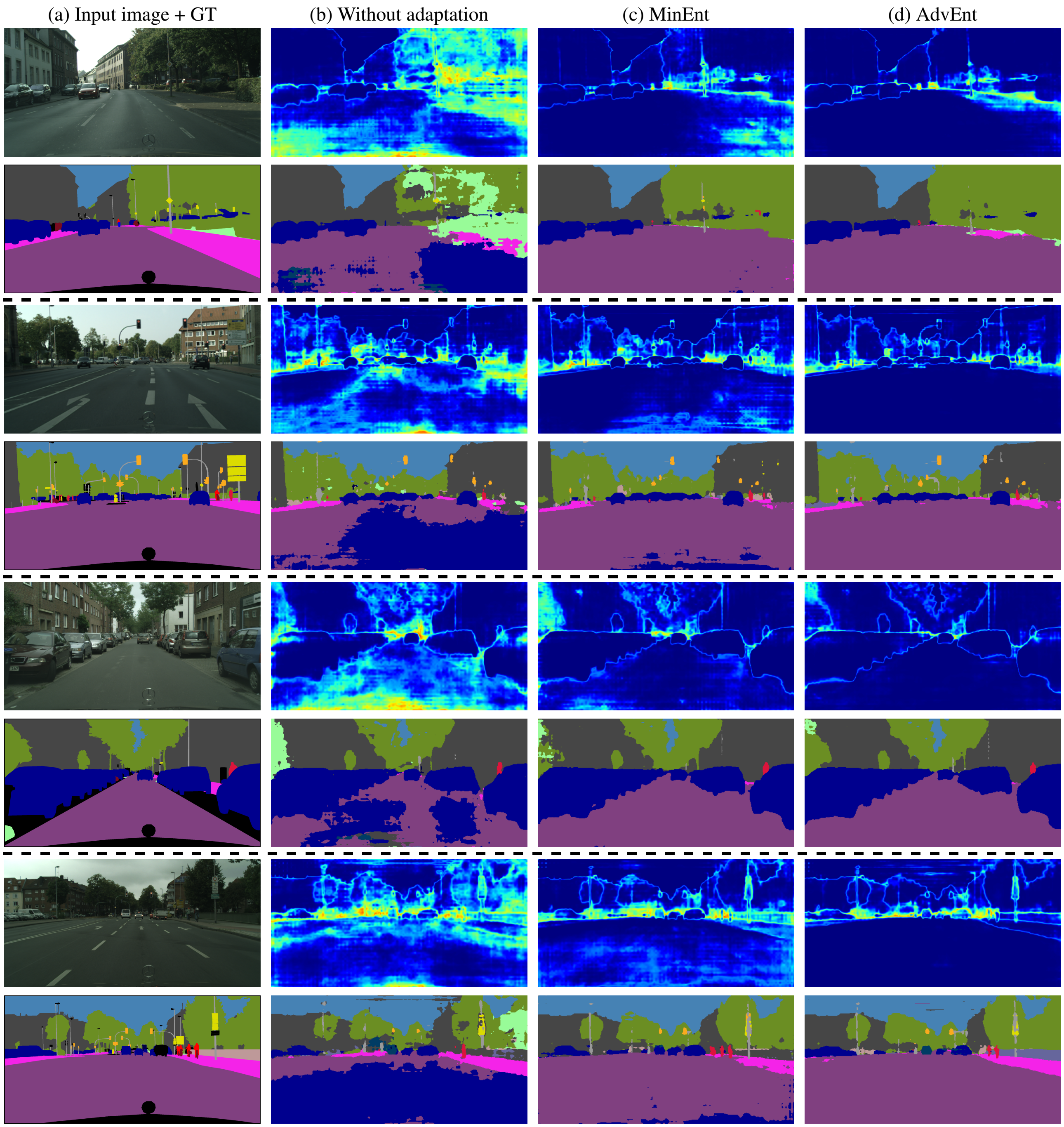

Figure 1: **Qualitative results of semantic segmentation**. Column (a) shows input images and the corresponding semantic segmentation ground-truths. Columns (b), (c) and (d) show segmentation results (upper) along with prediction entropy maps produced by different approaches (lower). For a better illustration, we provide a demo video in the archive.

Figure 2: **Qualitative results for object detection**. Column (a) shows the input images. Columns (b), (c) and (d) illustrate detection results of the baseline and of our MinEnt and AdvEnt models. Detections of different classes are plotted in different colors. We visualize all the detections with scores greater than 0.5.