Fully Test-time Adaptation for Object Detection Supplementary Materials

Xiaoqian Ruan and Wei Tang
University of Illonis Chicago
{xruan9,tangw}@uic.edu

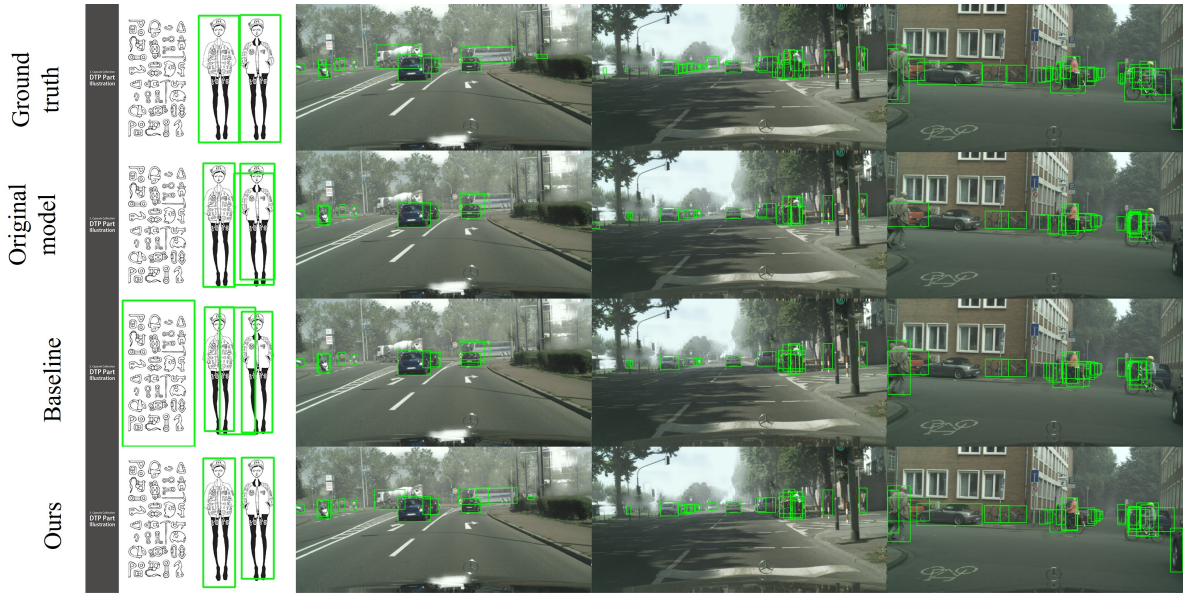

Figure 1. Qualitative results of fully test-time adaptation for object detection. The first row is the ground truth of each image, the second row is the results of the pre-trained detector, the third row is the results of the self-training baseline, and the last row is the results of our proposed method. All the object detections are denoted as green boxes.

1. Qualitative Results

Additional qualitative results are shown in Fig. 1. The first row is the ground truth of each sample, the second row is the object detections obtained by the original Faster RCNN[1] model, the third row is the results of the self-training baseline, and the last row is our proposed method's results. It demonstrates that our proposed method could remove the wrong detections and detect the missing objects effectively.

2. Self-training Framework

The pipeline of the self-training framework for fully test-time adaptation (FTTA) is shown in Algorithm 1. It is an iterative algorithm. At the tth iteration ($t \in \{1, \dots, T\}$), the current detector $f_{\boldsymbol{\theta}_{t-1}}$ makes a prediction $\mathcal{D}_t$ on the single testing image $\boldsymbol{I}$. $\boldsymbol{\theta}_{t-1}$ is the parameters of the detector. We then collect pseudo labels $\mathcal{P}_t$ from $\mathcal{D}_t$ based on the detection confidence (baseline) or our proposed IoU Filter. Finally, we tune the current parameters $\boldsymbol{\theta}_{t-1}$ on the pseudo labels via a gradient descent step and obtain the updated parameters $\boldsymbol{\theta}_t$. The loss L is the same as that used when the model was originally trained on the source data, except that the RoI classification loss is removed for better performance. At the first iteration, *i.e.*, $t = 1$, the parameters $\boldsymbol{\theta}_{t-1}$ are initialized as $\boldsymbol{\theta}_0$ obtained on the source data. After the last iteration, *i.e.*, $t = T$, the detector $f_{\boldsymbol{\theta}_T}$ will be used to make a final prediction on $\boldsymbol{I}$.

References

[1] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information process-*

Algorithm 1 Self-training Framework for FTTA

Input: A pre-trained detector $f_{\boldsymbol{\theta}_0}$, a single testing image $\boldsymbol{I}$, and the number of self-training iterations T

Output: The prediction $\mathcal{D}_T$

for $t = 1, ..., T$ **do**
    $\mathcal{D}_t \leftarrow f_{\boldsymbol{\theta}_{t-1}}(\boldsymbol{I})$
    $\mathcal{P}_t \leftarrow \texttt{FilterPseudoLabels}(\mathcal{D}_t)$
    $\boldsymbol{\theta}_t \leftarrow \boldsymbol{\theta}_{t-1} - \nabla_{\boldsymbol{\theta}_{t-1}} L(\mathcal{D}_t, \mathcal{P}_t)$
end for
$\mathcal{D}_T \leftarrow f_{\boldsymbol{\theta}_T}(\boldsymbol{I})$

ing systems, 28, 2015. 1