Supplementary Material for Self-Training and Adversarial Background Regularization for Unsupervised Domain Adaptive One-Stage Object Detection

Seunghyeon Kim
KAIST

Jaehoon Choi
KAIST

Taekyung Kim
KAIST

Changick Kim
KAIST

{seunghyeonkim, whdns44, tkkim93, changick}@kaist.ac.kr

1. More Ablation Study on BSR

Following [2], we defined α-balanced variant of background score regularization term as

$$L_{\alpha-adv}(x^t) = -t\sum_i \alpha|t - p_i(0|x^t)|^\gamma \cdot \log(p_i(0|x^t)) - (1-t)\sum_i \alpha|t - p_i(0|x^t)|^\gamma \cdot \log(1 - p_i(0|x^t)), \quad (1)$$

where α is a hyperparameter. α also can be shown as a parameter that balancing between the task loss and the adversarial loss.

$$\min_F L_{task}(x^s, y^s) - \alpha L_{adv}(x^t), \quad (2)$$

$$\min_C L_{task}(x^s, y^s) + \alpha L_{adv}(x^t), \quad (3)$$

Thus, the base network is identical to when the α is set to 0. We used $\alpha = 1.0$ in the main paper.

We further present an ablation study on the hyperparameter $\alpha \in \{0.0, 0.2, 0.5, 1.0, 1.5, 2.0\}$ with $\gamma = 2.0$ and $t = 0.5$ in Eq. (1). As shown in Fig. 1, the network shows poor performance gain with small value of α since the proposed adversarial background score regularization (BSR) loss L_{adv} contributes less. We experimentally found that the network shows the best performance with $\alpha = 1.0$. On the other hand, the network can be over-regularized with large value of α. Though, the performance is not curcially relied on the value of α around 1.0.

2. Error Analysis

We also progressed the error analysis following [1]. We provide the pie chart of false positives of the person class and the furniture classes in Fig. 2 and 3. From the results, we found that the performance of the person class is improved with large reduction of false negatives.

Interestingly, WST reduces localization error even it does not containing localization loss. From this result, we insist that only enhancing classification can also strengthen the localization.

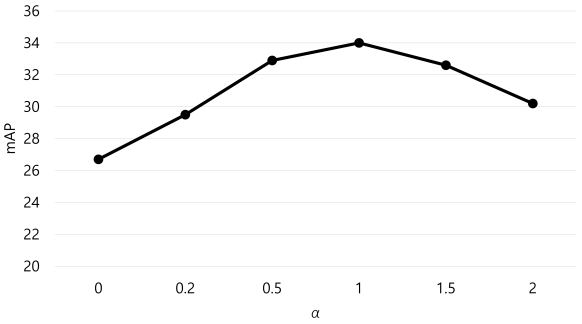

Figure 1. Trends of mAP with various values of α.

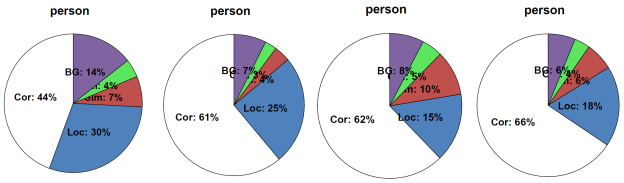

Figure 2. Error analysis of the person class. From the left, the base network, BSR, WST, and both.

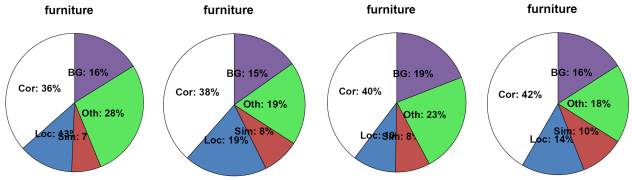

Figure 3. Error analysis of the furniture classes. From the left, the base network, BSR, WST, and both.

3. More Qualitative Results

Figure 4 shows more qualitative results of the base network, the proposed method, and the ground-truth from left to right. We present images containing many objects.

Figure 4. Qualitative results of images containing many objects.

References

[1] Derek Hoiem, Yodsawalai Chodpathumwan, and Qieyun Dai. Diagnosing error in object detectors. In *European conference on computer vision*, pages 340–353. Springer, 2012. 1

[2] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 1