Supplementary Material: Informative and Consistent Correspondence Mining for Cross-Domain Weakly Supervised Object Detection

1. Derivation of Approximated Objective

In the Eqn. (6) of our submission, we propose to use the following factorizable posterior

$$Q(\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_-}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_{\mathcal{R}}},\mathbb{I}) = \int P(\mathbf{a}_{\mathcal{R}}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_{\mathcal{R}}},\mathbb{I})P(\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_-}|\mathbf{a}_{\mathcal{R}},\mathbb{I})d\mathbf{a}_{\mathcal{R}}, \tag{1}$$

to approximate the original posterior $P(\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_-}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_{\mathcal{R}}},\mathbb{I})$ when computing its entropy. Detailed derivation is as follows.

First we note that for a specific region $\mathcal{R}$, its annotation $\mathbf{a}_{\mathcal{R}}$ is deterministic thus $P(\mathbf{a}|\mathbb{I})$ is a delta distribution in that it takes 1 when $\mathbf{a} = \mathbf{a}_{\mathcal{R}}$ and 0 otherwise. Therefore, $P\left(\mathbf{a}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_+},\mathbb{I}\right) = \frac{P\left(\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_+}|\mathbf{a},\mathbb{I}\right)P(\mathbf{a}|\mathbb{I})}{P\left(\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_+}|\mathbb{I}\right)}$ is also a delta distribution. Thus, the posterior (1) is only determined by its mode, or Maximum-A-Posterior (MAP) solution of $P\left(\mathbf{a}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_+},\mathbb{I}\right)$

$$\begin{aligned} Q\left(\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_-}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}^+},\mathbb{I}\right) &= P\left(\mathbf{a}_{\mathcal{R}}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_-},\mathbb{I}\right)P\left(\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_-}|\mathbb{I}\right) \\ \text{s.t. } \mathbf{a}_{\mathcal{R}} &= \arg\max_{\mathbf{a}} P\left(\mathbf{a}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_+},\mathbb{I}\right). \end{aligned} \tag{2}$$

Then, using the rule of conditional entropy computation,

$$\begin{aligned} &H\left(\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_-}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_+},\mathbb{I}\right) \approx H_Q\left(\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_-}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_+},\mathbb{I}\right) \\ &= H_Q\left(\mathbf{a}_{\mathcal{R}}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_-},\mathbb{I}\right) + H_Q\left(\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_-}|\mathbb{I}\right) - H_Q\left(\mathbf{a}_{\mathcal{R}}|\mathbb{I}\right), \end{aligned} \tag{3}$$

Since that $H\left(\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_-}|\mathbb{I}\right)$ and $H\left(\mathbf{a}_{\mathcal{R}}|\mathbb{I}\right)$ are constants, they are eliminated from the objective. The original objective Eqn. (3) in our submission therefore writes as

$$\begin{aligned} \max_{\mathbf{\Omega}} A\left(\mathcal{S},\mathcal{T}\right) &= \frac{1}{Z_A}\sum_{\mathbb{I}}\sum_{\mathcal{R}\in\mathbb{R}_{\mathcal{S}}}\sum_{\mathcal{C}_-} H\left(\mathbf{a}_{\mathcal{R}}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_-},\mathbb{I}\right), \\ \text{s.t. } \mathbf{a}_{\mathcal{R}} &= \arg\max_{\mathbf{a}} P\left(\mathbf{a}|\mathbf{w}_{\mathcal{R}}^{\mathcal{C}_{\mathcal{R}}},\mathbb{I}\right), \end{aligned} \tag{4}$$

where Z_A is a normalization constant. This equals to Eqn. (6) of our submission.

2. More Visual Results

In Fig. 1 $\sim$ Fig. 10, we show more results generated by different approaches, including STABR [3], SWDA [4], HTD [1], CSWSDA [2] and the proposed approach.

References

[1] C.i Chen, Z. Zheng, X. Ding, Y. Huang, and Q. Dou. Harmonizing transferability and discriminability for adapting object detectors. In *CVPR*, pages 8866–8875, 2020. 1

[2] N. Inoue, R. Furuta, T. Yamasaki, and K. Aizawa. Cross-domain weakly-supervised object detection through progressive domain adaptation. In *CVPR*, pages 5001–5009, 2018. 1

[3] S. Kim, J. Choi, T. Kim, and C. Kim. Self-training and adversarial background regularization for unsupervised domain adaptive one-stage object detection. In *ICCV*, pages 6091–6100, 2019. 1

[4] K. Saito, Y. Ushiku, T. Harada, and K. Saenko. Strong-weak distribution alignment for adaptive object detection. In *CVPR*, pages 6956–6965, 2019. 1

Figure 1. Representative results generated by different approaches (visualized in different rows). Best viewed with zoom in.

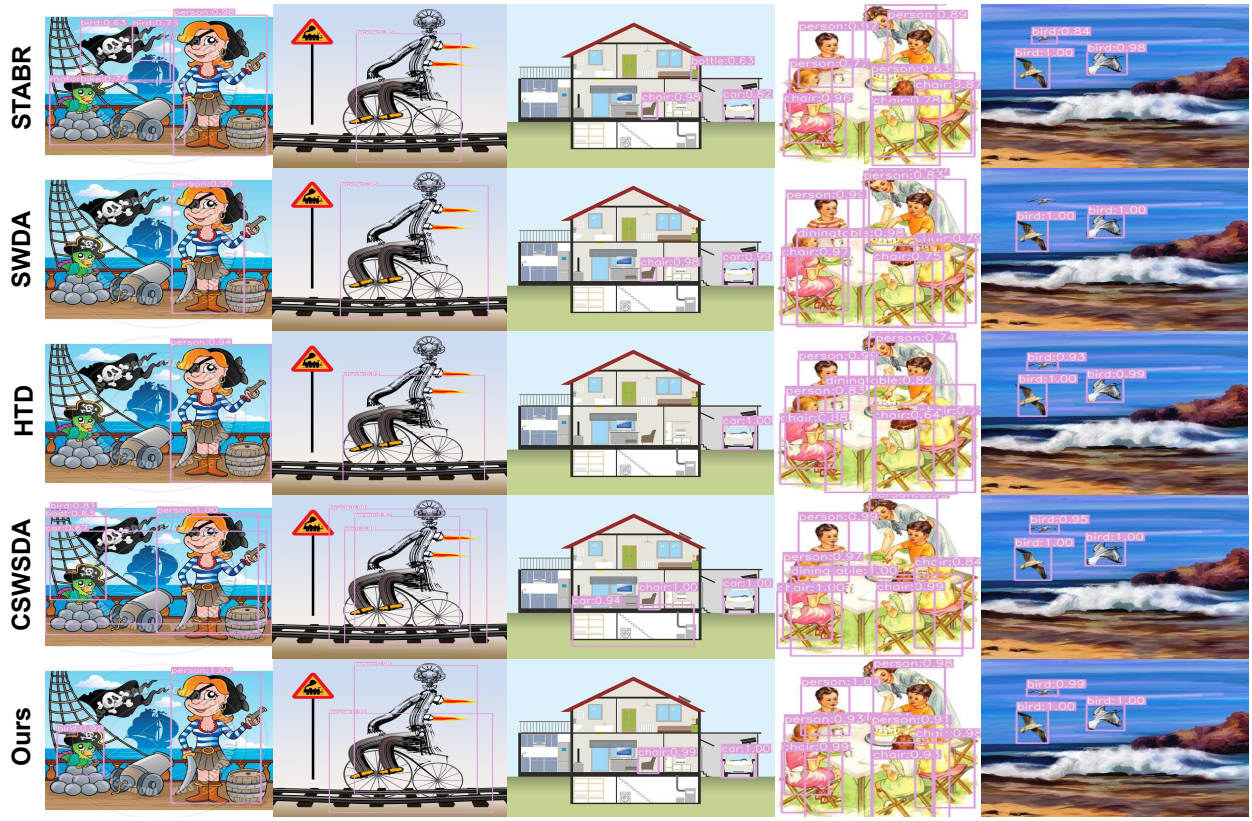

Figure 2. Representative results generated by different approaches (visualized in different rows). Best viewed with zoom in.

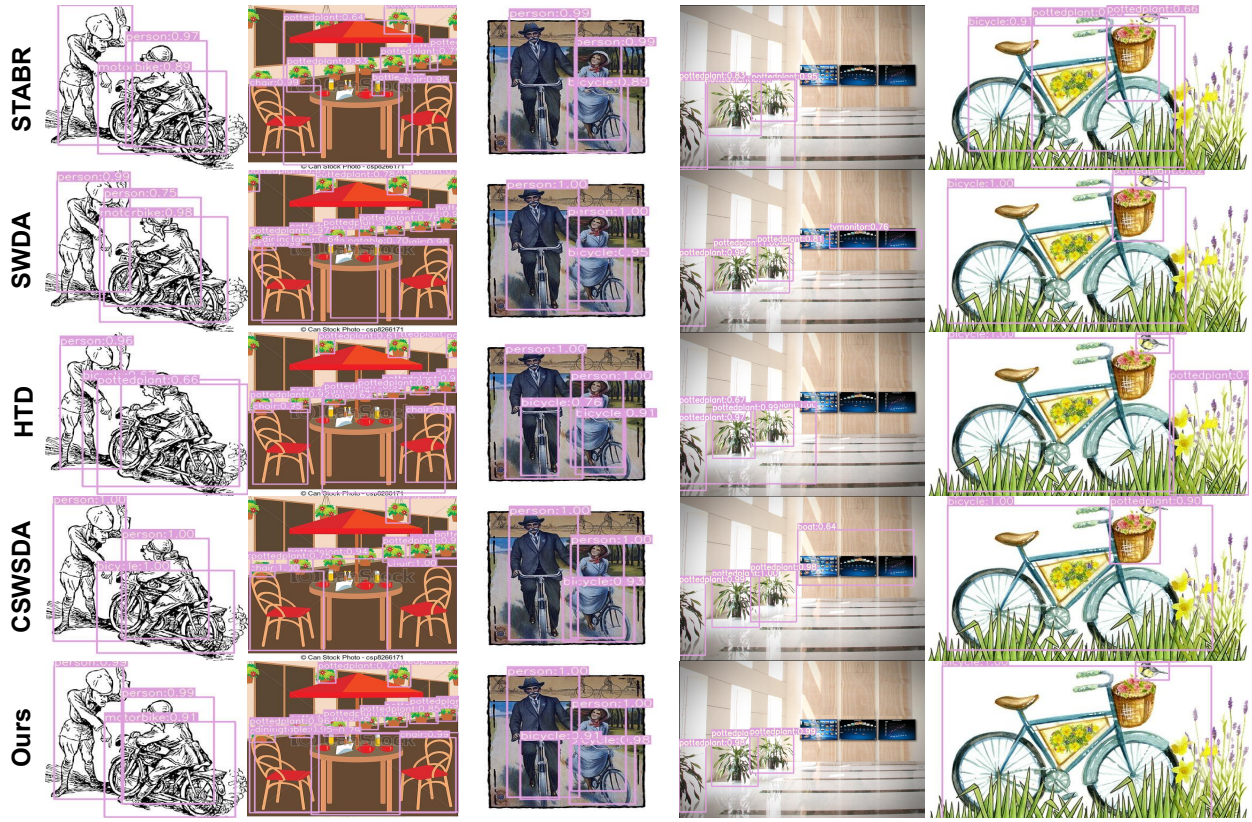

Figure 3. Representative results generated by different approaches (visualized in different rows). Best viewed with zoom in.

Figure 4. Representative results generated by different approaches (visualized in different rows). Best viewed with zoom in.

Figure 5. Representative results generated by different approaches (visualized in different rows). Best viewed with zoom in.

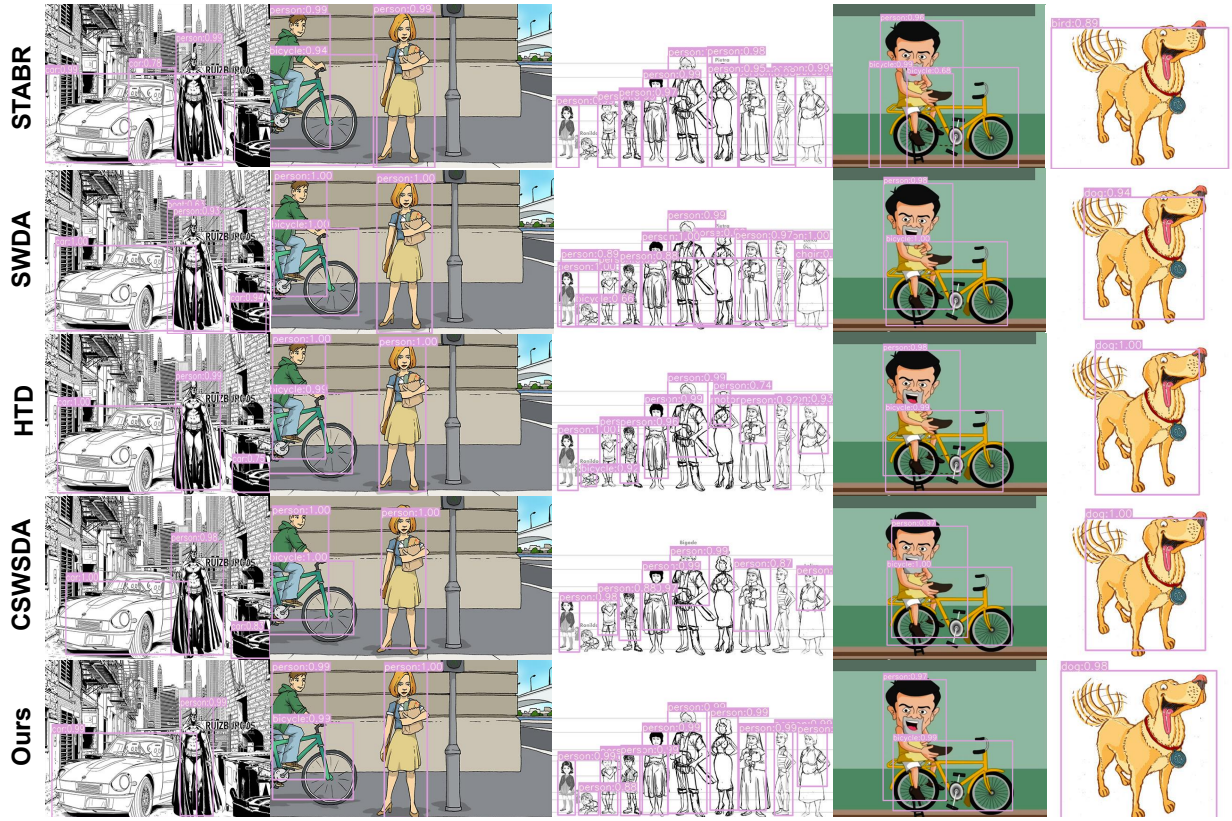

Figure 6. Representative results generated by different approaches (visualized in different rows). Best viewed with zoom in.

Figure 7. Representative results generated by different approaches (visualized in different rows). Best viewed with zoom in.

Figure 8. Representative results generated by different approaches (visualized in different rows). Best viewed with zoom in.

Figure 9. Representative results generated by different approaches (visualized in different rows). Best viewed with zoom in.

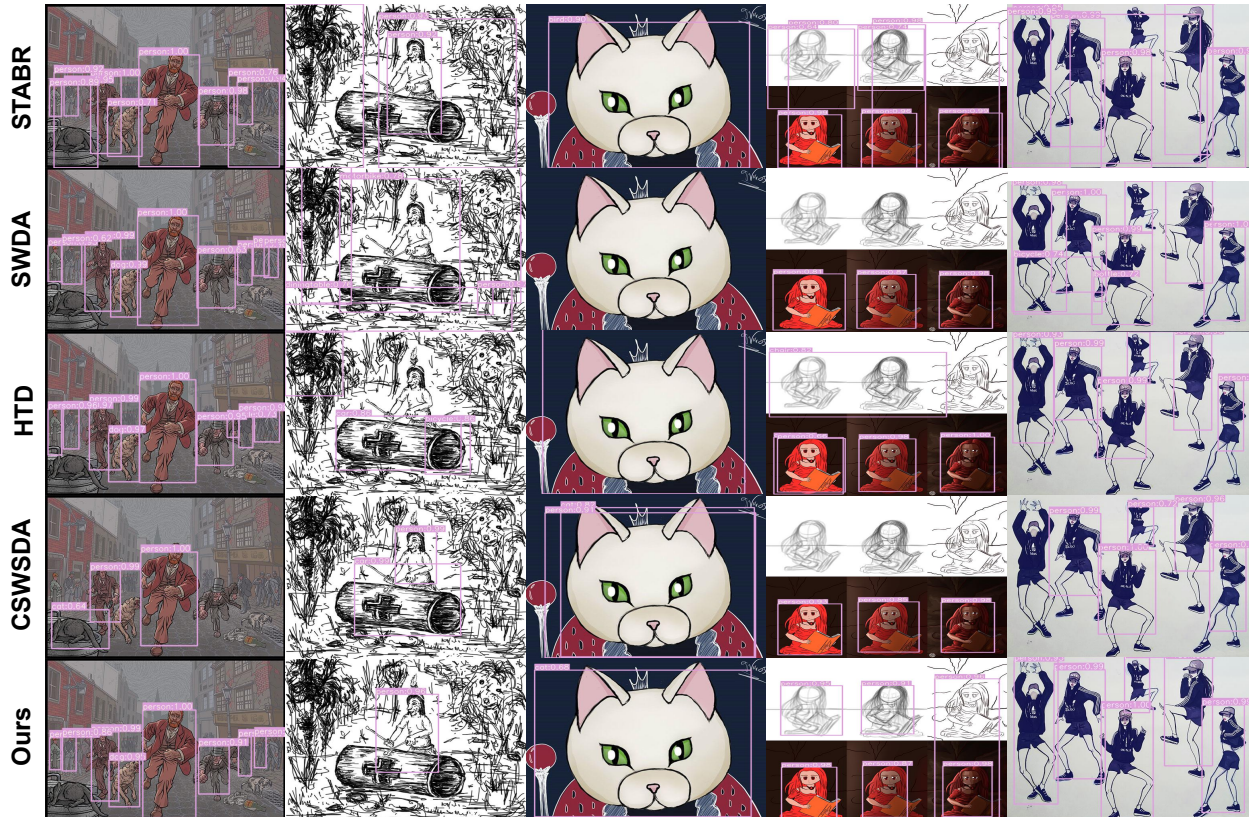

Figure 10. Representative results generated by different approaches (visualized in different rows). Best viewed with zoom in.