Domain Adaptation for Semantic Segmentation with Maximum Squares Loss -Supplementary Material-

Minghao Chen, Hongyang Xue, Deng Cai*

State Key Lab of CAD&CG, College of Computer Science, Zhejiang University, Hangzhou, China
Fabu Inc., Hangzhou, China
Alibaba-Zhejiang University Joint Institute of Frontier Technologies, Hangzhou, China
minghaochen01@gmail.com, hyxue@outlook.com, dengcai@cad.zju.edu.cn

A. Derivation for Relation with f-divergence

In Section 3.2.2, we explain the maximum squares loss from f-divergence view.

In Eq.10, we consider maximizing the Person χ^2 divergence ($f(t) = x^2 - 1$) between the prediction distribution p and the uniform distribution $\mathcal{U}$, which is equivalent to minimizing our maximum square loss: $-\sum_c \left(p_{x_t}^{n,c}\right)^2$. The derivation of Eq.10 is:

$$\begin{aligned} D_{\chi^2}(p\|\mathcal{U}) &= \sum_c \frac{1}{C}(C^2 p_{x_t}^{n,c2} - 1)) \\ &= \sum_c (C p_t^{n,c2}) - \sum_c \frac{1}{C} \\ &= C\sum_c \left(p_t^{n,c}\right)^2 - 1. \end{aligned}$$

Meanwhile, we show that minimizing the entropy H is equivalent to maximizing the KL-divergence ($f = t\log t$) between the prediction distribution p and the uniform distribution $\mathcal{U}$, of which each class has the same probability $\frac{1}{C}$. The derivation is :

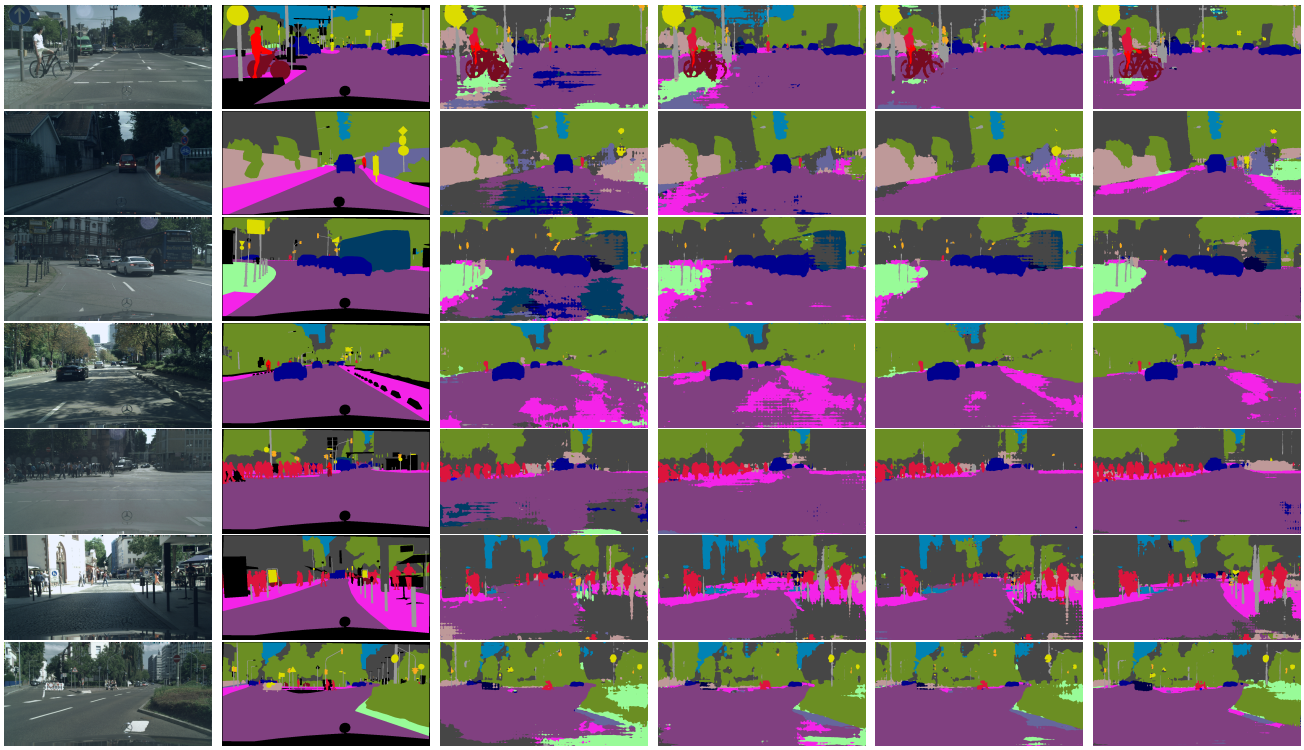
$$\begin{aligned} D_{KL}(p_t^{n,c}\|\mathcal{U}) &= \sum_c (p_t^{n,c}\log(C) + p_t^{n,c}\log(p_t^{n,c})) \\ &= \sum_c (p_t^{n,c}\log(C)) + \sum_c (p_t^{n,c}\log(p_t^{n,c})) \\ &= \sum_c (p_t^{n,c}\log(p_t^{n,c})) + \log(C) \\ &= -H(p_t^{n,c}) + \log(C). \end{aligned}$$

B. Example Results of GTA5-to-Cityscapes

In Section 4.3, we set GTA5 as the labeled source domain and Cityscapes as the unlabeled target domain. We show our maximum squares loss, "MaxSquare", exceeds the original entropy minimization loss by a large margin. Moreover, combined with our image-wise weighting factor and the multi-level self-produced guidance, our method "MaxSquare+IW+Multi" achieves state-of-the-art performance. The following Figure 1 shows the example results of GTA5-to-Cityscapes experiment with ResNet-101 backbone.

As Figure 1 demonstrates, compared with the entropy minimization method "MinEnt", our maximum squares loss "MaxSquare" can handle some hard-to-classify areas, such as some road areas are easy to be mistaken for the sidewalk class by "MinEnt". Furthermore, combined with the image-wise weighting factor "IW", the model can predict more accurate on the small class, such as the person and the bike class.

*Corresponding author



(a) Image (b) GT (c) Source only (d) MinEnt (e) MaxSquare (f) MaxSquare+IW

Figure 1: (a) The image in Cityscapes. (b) The ground truth. (c) Directly transfer the model trained on GTA5. (d) Domain adaptation with the entropy minimization method. (e) Domain adaptation with the maximum squares loss. (f) Domain adaptation using the maximum squares loss combined with the image-wise weighting factor.