Progressive Feature Alignment for Unsupervised Domain Adaptation Supplementary Material

Chaoqi Chen*[1], Weiping Xie*[1], Wenbing Huang[2], Yu Rong[2], Xinghao Ding[1],
Yue Huang†[1], Tingyang Xu†[2], Junzhou Huang[2]
[1] Fujian Key Laboratory of Sensing and Computing for Smart City,
School of Information Science and Engineering, Xiamen University, China
[2] Tencent AI Lab
cqchen94@stu.xmu.edu.cn, xiewp@stu.xmu.edu.cn, hwenbing@126.com, yu.rong@hotmail.com
dxh@xmu.edu.cn, huangyue05@gmail.com, Tingyangxu@tencent.com, jzhuang@uta.edu

1. Proof of Theorem

We resort to the pseudo-labels to bound the combined error of the ideal hypothesis, C. Then, the following inequality holds:

Theorem 1. *Let $f_{\hat{\mathcal{T}}}$ be the pseudo-labeling function. Given $R_{\mathcal{T}'}(f_{\mathcal{S}}, f_{\hat{\mathcal{T}}})$ and $R_{\mathcal{T}'}(f_{\mathcal{T}}, f_{\hat{\mathcal{T}}})$ as the minimum shared error and the degree to which the target samples are falsely labeled on $\hat{D}_t$, respectively. We have*

$$C \leq \min_{h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}'}(h, f_{\hat{\mathcal{T}}}) + 2R_{\mathcal{T}'}(f_{\mathcal{S}}, f_{\hat{\mathcal{T}}}) + R_{\mathcal{T}'}(f_{\mathcal{T}}, f_{\hat{\mathcal{T}}}). \quad (1)$$

Proof. This proof relies on the triangle inequality for classification error [1, 2], which implies that for any labeling functions f_1, f_2, and f_3, we have $R(f_1, f_2) \leq R(f_1, f_3) + R(f_2, f_3)$. Then

$$\begin{aligned} C &= \min_{\forall h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}'}(h, f_{\mathcal{T}}) \\ &\leq \min_{\forall h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}'}(h, f_{\mathcal{S}}) + R_{\mathcal{T}'}(f_{\mathcal{S}}, f_{\mathcal{T}}) \\ &\leq \min_{\forall h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}'}(h, f_{\hat{\mathcal{T}}}) + R_{\mathcal{T}'}(f_{\mathcal{S}}, f_{\hat{\mathcal{T}}}) \\ &+ R_{\mathcal{T}'}(f_{\mathcal{S}}, f_{\hat{\mathcal{T}}}) + R_{\mathcal{T}'}(f_{\mathcal{T}}, f_{\hat{\mathcal{T}}}) \\ &= \min_{\forall h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}'}(h, f_{\hat{\mathcal{T}}}) \\ &+ 2R_{\mathcal{T}'}(f_{\mathcal{S}}, f_{\hat{\mathcal{T}}}) + R_{\mathcal{T}'}(f_{\mathcal{T}}, f_{\hat{\mathcal{T}}}) \end{aligned}$$

Therefore, the main inequality (1) holds. □

2. Experiments

More Training Details. For Office-31 and ImageCLEF-DA datasets (AlexNet-based): (1) we augment the input images by scaling all images to 256×256, randomly cropping 227×227 patches and executing random flips; (2) we set the batch size to 100 for each domain; (3) We add a bottleneck layer fcb with 256 units after the fc7 layer for safer transfer representation learning; (4) the discriminator consists of 3 fully connected layers: $1024 \rightarrow 1024 \rightarrow 1$. For digit recognition datasets: (1) all images are cast to 32×32×1 in all experiments for fair comparison; (2) we set the batch size to 128 for each domain; (3) the discriminator consists of 3 fully connected layers: $500 \rightarrow 500 \rightarrow 1$; (4) We use the Adam optimizer with learning rates set to 0.01 instead of the learning rate annealing strategy.

* indicates equal contributions.
† Corresponding authors

Table 5: Comparing Different Temperature T on A→W.

T	1.5	1.6	1.7	1.8	1.9	2.0
Accuracy (%)	80.9	81.8	82.1	**83.0**	80.9	80.7

Comparing Different Temperature T. We perform sensitivity analysis of the temperature T on transfer task **A** → **W**. We provide the classification accuracy as the changing of T in [1.5, 0.1, 2.0] and the results are reported in Table 5. The accuracy first increases and then decreases as T varies and shows a best result when $T = 1.8$. The results implicitly confirm our assumption that a good UDA model needs a non-saturated source classifier.

References

[1] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. In *Advances in neural information processing systems*, pages 137–144, 2007.

[2] Koby Crammer, Michael Kearns, and Jennifer Wortman. Learning from multiple sources. *Journal of Machine Learning Research*, 9(Aug):1757–1774, 2008.