Marginalized Augmented Few-shot Domain Adaptation - Supplementary

I. Proof of Proposition I

Proposition I *Given synthesized samples $\tilde{\mathcal{D}}_{SSA} \in \mathbb{R}^{MKC}$, as $M/K \to \infty$, the expected cross-entropy loss $\mathcal{L}_{aug}^{\infty}$ is upper-bounded as $\overline{\mathcal{L}}_{aug}^{\infty}$, which can be calculated as follows:*

$$\begin{aligned}\mathcal{L}_{aug}^{\infty} &= \mathbb{E}_c \mathbb{E}_{\hat{\boldsymbol{\mu}}^c} \mathbb{E}_{\tilde{\mathbf{f}}_k^c}\Big[-\log\big(\frac{e^{\mathbf{w}_c^\top \tilde{\mathbf{f}}_k^c + b_c}}{\sum_{j=1}^C e^{\mathbf{w}_j^\top \tilde{\mathbf{f}}_k^c + b_j}}\big)\Big] \\ &\leq \mathbb{E}_c\Big[-\log\Big(\frac{e^{\mathbf{w}_c^\top((1-\beta)\boldsymbol{\eta}_s^c + \beta\boldsymbol{\eta}_t^c) + b_c}}{\sum_{j=1}^C e^{\mathbf{w}_j^\top((1-\beta)\boldsymbol{\eta}_s^c + \beta\boldsymbol{\eta}_t^c) + b_j + \mathcal{A}}}\Big)\Big],\end{aligned} \tag{1}$$

where $\mathcal{A} = \frac{\alpha}{2}(\mathbf{w}_j^\top - \mathbf{w}_c^\top)\boldsymbol{\Sigma}_s^c(\mathbf{w}_j - \mathbf{w}_c)$.

Proof. For k^{th} *anchor* $\hat{\boldsymbol{\mu}}^{c(k)}$ of class c, the augmented samples by SSA are $\tilde{\mathcal{D}}_{\mathbf{SSA}}^{c(k)} = (\tilde{\mathbf{f}}_k^{c(1)}, c), \cdots, (\tilde{\mathbf{f}}_k^{c(M)}, c)\}$ of size M, where $\tilde{\mathbf{f}}_k^{c(m)}$ is the m^{th} augmented feature given the synthesized intermediate *anchor* $\hat{\boldsymbol{\mu}}^{c(k)}$. Then the expected cross-entropy loss is defined as:

$$\begin{aligned}\lim_{M\to\infty} \mathcal{L}_{aug}^{c(k)} &= \frac{1}{M}\sum_{m=1}^{M} -\log\Big(\frac{e^{\mathbf{w}_c^\top \tilde{\mathbf{f}}_k^{c(m)} + b_c}}{\sum_{j=1}^C e^{\mathbf{w}_j^\top \tilde{\mathbf{f}}_k^{c(m)} + b_j}}\Big) \\ &= \mathbb{E}_{\tilde{\mathbf{f}}_k^c}\Big[-\log\Big(\frac{e^{\mathbf{w}_c^\top \tilde{\mathbf{f}}_k^c + b_c}}{\sum_{j=1}^C e^{\mathbf{w}_j^\top \tilde{\mathbf{f}}_k^c + b_j}}\Big)\Big] \\ &= \mathbb{E}_{\tilde{\mathbf{f}}_k^c}\Big[\log\big(\sum_{j=1}^C e^{(\mathbf{w}_j^\top - \mathbf{w}_c^\top)\tilde{\mathbf{f}}_k^c + (b_j - b_c)}\big)\Big] \\ &\leq \log\Big(\sum_{j=1}^C \mathbb{E}_{\tilde{\mathbf{f}}_k^c}\Big[e^{(\mathbf{w}_j^\top - \mathbf{w}_c^\top)\tilde{\mathbf{f}}_k^c + (b_j - b_c)}\Big]\Big),\end{aligned} \tag{2}$$

where inequality follows the Jensen's inequality $\mathbf{E}[\log(X)] \leq \log(\mathbf{E}[X])$ [1], as the logarithmic function $\log(\cdot)$ is concave. The upper-bound of $\lim_{M\to\infty} \mathcal{L}_{aug}^{c(k)}$ is obtained by leveraging the moment-generating function $M_X(t) = \mathbf{E}(e^{tX}), t \in \mathbb{R}$. Specifically, for $\tilde{\mathbf{f}}_k^{c(m)} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}^{c(k)}, \alpha\Sigma_s^c)$ which is drawn from a Gaussian distribution, it is provable that $(\mathbf{w}_j^\top - \mathbf{w}_c^\top)\tilde{\mathbf{f}}_k^{c(m)} + (b_j - b_c)$ follows Gaussian distribution, i.e., $(\mathbf{w}_j^\top - \mathbf{w}_c^\top)\tilde{\mathbf{f}}_k^{c(m)} + (b_j - b_c) \sim \mathcal{N}((\mathbf{w}_j^\top - \mathbf{w}_c^\top)\hat{\boldsymbol{\mu}}^{c(k)} + (b_j - b_c), \alpha(\mathbf{w}_j^\top - \mathbf{w}_c^\top)\Sigma_s^c((\mathbf{w}_j - \mathbf{w}_c)))$. Referring to the moment-generating function of Gaussian distribution: $\mathbf{E}[e^{tX}] = e^{t\mu + \frac{1}{2}\sigma^2 t^2}, X \sim \mathcal{N}(\mu, \sigma^2)$, we have the upper bound $\lim_{M\to\infty} \mathcal{L}_{aug}^{c(k)}$ as:

$$\lim_{M\to\infty} \mathcal{L}_{aug}^{c(k)} \leq -\log\Big(\frac{e^{\mathbf{w}_c^\top \hat{\boldsymbol{\mu}}^{c(k)} + b_c}}{\sum_{j=1}^C e^{\mathbf{w}_j^\top \hat{\boldsymbol{\mu}}^{c(k)} + b_j + \mathcal{A}}}\Big), \tag{3}$$

where $\mathcal{A} = \frac{\alpha}{2}(\mathbf{w}_j^\top - \mathbf{w}_c^\top)\boldsymbol{\Sigma}_s^c(\mathbf{w}_j - \mathbf{w}_c)$. Moreover, as there are K synthesized intermediate *anchor* $\hat{\boldsymbol{\mu}}^{c(k)}$ generated by CCA, the overall expected cross-entropy loss for all augmented samples based on all possible *anchors* are:

$$\begin{aligned}\mathcal{L}_{aug}^{\infty} &= \lim_{\substack{M\to\infty \\ K\to\infty}} \mathbb{E}_c\Big[\frac{1}{K}\sum_{k=1}^{K} \mathcal{L}_{aug}^{c(k)}\Big] \\ &= \mathbb{E}_c \mathbb{E}_{\hat{\boldsymbol{\mu}}^c}\Big[-\log\Big(\frac{e^{\mathbf{w}_c^\top \hat{\boldsymbol{\mu}}^{c(k)} + b_c}}{\sum_{j=1}^C e^{\mathbf{w}_j^\top \hat{\boldsymbol{\mu}}^{c(k)} + b_j + \mathcal{A}}}\Big)\Big] \\ &\leq \mathbb{E}_c\Big[\log\Big(\sum_{j=1}^C \mathbb{E}_{\hat{\boldsymbol{\mu}}^c}\big[e^{(\mathbf{w}_j^\top - \mathbf{w}_c^\top)\hat{\boldsymbol{\mu}}^c + (b_j - b_c) + \mathcal{A}}\big]\Big)\Big].\end{aligned} \tag{4}$$

As we know that $\hat{\boldsymbol{\mu}}^c = (1-\lambda\beta)\boldsymbol{\eta}_s^c + \lambda\beta\boldsymbol{\eta}_t^c = \beta(\boldsymbol{\eta}_t^c - \boldsymbol{\eta}_s^c)\lambda + \boldsymbol{\eta}_s^c$, and $\lambda \sim Beta(a, b)$ follows Beta distribution. Thus,

$$\begin{aligned}\mathcal{L}_{aug}^{\infty} &\leq \mathbb{E}_c\Big[\log\Big(\sum_{j=1}^C \mathbb{E}_{\hat{\boldsymbol{\mu}}^c}\big[e^{(\mathbf{w}_j^\top - \mathbf{w}_c^\top)\hat{\boldsymbol{\mu}}^c + (b_j - b_c) + \mathcal{A}}\big]\Big)\Big] \\ &= \mathbb{E}_c\Big[\log\Big(\sum_{j=1}^C \mathbb{E}_{\lambda}\big[e^{\beta(\mathbf{w}_j^\top - \mathbf{w}_c^\top)(\boldsymbol{\eta}_t^c - \boldsymbol{\eta}_s^c)\lambda}\big]e^{\mathcal{A} + \mathcal{B}}\Big)\Big],\end{aligned} \tag{5}$$

where $\mathcal{A} = \frac{\alpha}{2}(\mathbf{w}_j^\top - \mathbf{w}_c^\top)\boldsymbol{\Sigma}_s^c(\mathbf{w}_j - \mathbf{w}_c)$, $\mathcal{B} = (\mathbf{w}_j^\top - \mathbf{w}_c^\top)\boldsymbol{\eta}_s^c + (b_j - b_c)$.

As the moment-generating function of Beta distribution is defined as: $\mathbf{E}[e^{tX}] = 1 + \sum_{k=1}^{\infty}(\prod_{r=0}^{k-1}\frac{a+r}{a+b+r})\frac{t^k}{k!}, X \sim \mathrm{Beta}(a, b)$. and $a, b > 0$, such that $\prod_{r=0}^{k-1}\frac{a+r}{a+b+r} < 1$, then we obtain $\mathbf{E}[e^{tX}] \leq 1 + \sum_{k=1}^{\infty}\frac{t^k}{k!} = e^t$, thus the upper bound of $\mathcal{L}_{aug}^{\infty}$ is obtained as:

$$\mathcal{L}_{aug}^{\infty} \leq \mathbb{E}_c\Big[-\log\Big(\frac{e^{\mathbf{w}_c^\top((1-\beta)\boldsymbol{\eta}_s^c + \beta\boldsymbol{\eta}_t^c) + b_c}}{\sum_{j=1}^C e^{\mathbf{w}_j^\top((1-\beta)\boldsymbol{\eta}_s^c + \beta\boldsymbol{\eta}_t^c) + b_j + \mathcal{A}}}\Big)\Big], \tag{6}$$

□

II. Ablation Analysis

A. Experimental Comparison of Framework Variants

To further demonstrate the effectiveness of the category-level domain confusion loss based on the prediction produced by the dual-classifier module, we compare the performance of the proposed model with several variants on task A D on the Office-31 dataset, and the results are reported in Figure 1. Specifically, Our proposed model is trained with adversarial domain discrimination and category-level confusion losses (Eq. (2)(3)(4) in the manuscript), which is denoted as "Ours." As a comparison, "w/o domain discriminator" denotes the results produced without any cross-domain domain alignment, only the classification supervision Cross-Entropy loss remains. In addition, "w/ feature alignment" denotes the results obtained via adding the MMD constraint to the features extracted from the backbone feature extractor, which is a widely used strategy to address domain adaptation problems [2]. Further, "w/

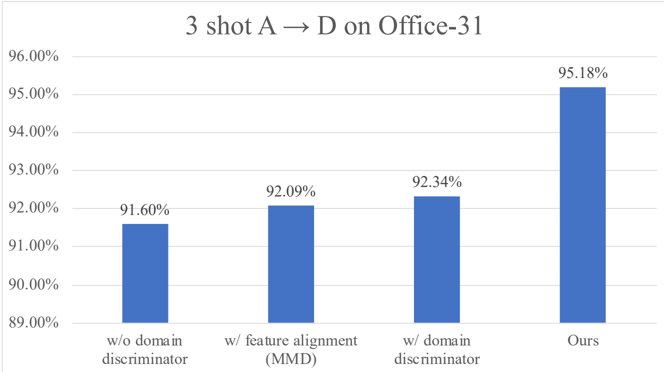

Fig. 1: Comparison of variants of the proposed framework on Office-31 dataset A D 3-shot task.

domain discriminator" shows the results that were obtained with a domain discriminator equipped as many unsupervised domain adaptation methods did [3]. The rest of the framework and training processes are identical to the whole framework. According to the results, we find significant performance boosts that demonstrate the contribution of category-level adversarial domain discrimination and confusion losses.

References

[1] J. L. W. V. Jensen *et al.*, "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," *Acta mathematica*, vol. 30, pp. 175–193, 1906.

[2] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," *arXiv preprint arXiv:1412.3474*, 2014.

[3] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7167–7176.