Supplemental File of BMD: A General Class-balanced Multicentric Dynamic Prototype Strategy for Source-free Domain Adaptation

Sanqing Qu[1], Guang Chen[1(✉)], Jing Zhang[2], Zhijun Li[3], Wei He[4], and Dacheng Tao[2,5]

[1] Tongji University {2011444, guangchen}@tongji.edu.cn
[2] The University of Sydney jing.zhang1@sydney.edu.au
[3] University of Science and Technology of China zjli@ieee.org
[4] University of Science and Technology Beijing weihe@ieee.org
[5] JD Explore Academy dacheng.tao@gmail.com

A Theoretical Understanding

As presented in [2], the expected error $\mathcal{E}_{\mathcal{T}}(h)$ of a hypothesis on the target domain can be decomposed into three items: (1) expected error of h on the source domain $\mathcal{E}_{\mathcal{S}}(h)$; (2) $\mathcal{H}\Delta\mathcal{H}$-distance $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S},\mathcal{T})$, regarding domain shift as the discrepancy between the disagreement of two hypothesis $h, h' \in \mathcal{H}\Delta\mathcal{H}$; and (3) the error λ of the ideal joint hypothesis on both source and target domains, $\lambda = \min_{h\in\mathcal{H}}[\mathcal{E}_{\mathcal{S}}(h) + \mathcal{E}_{\mathcal{T}}\mathcal{T}(h)]$. Formally, the error bound is:

$$\mathcal{E}_{\mathcal{T}}(h) \leq \mathcal{E}_{\mathcal{S}}(h) + \frac{1}{2}d_{\mathcal{H}\Delta\mathcal{H}} + \lambda \tag{1}$$

In Inequality 1, $\mathcal{E}_{\mathcal{S}}(h)$ is easily minimized by the supervised loss during source model pre-training. Since we cannot access to source domain data during model adaptation, learning a target-specific feature encoding module to generate target data representations that are well-aligned with source data representations is a promising direction [4,5]. Specifically, following [4,5] we fixed the source classifier and then introduce the feature prototype based pseudo-labeling strategy to realize target encoding module learning and mitigate the divergence $d_{\mathcal{H}\Delta\mathcal{H}}$ item. However, existing strategies [4,5,1] are implemented with instance-level prediction results, which are category-biased and tend to introducing noisy labels, since the visual domain gaps between source and target are usually different between categories [11,12]. To mitigate this pitfall, our BMD strategy introduces inter-class balanced sampling and intra-class multicentric prototype strategy, such that those hard-transfer data will not be dominated by those easy-transfer data, and we can thus minimize the $d_{\mathcal{H}\Delta\mathcal{H}}$ and λ.

B Experiments

B.1 Ablation study on Clustering Algorithm

In addition to applying the vanilla k-means algorithm for intra-class multicentric prototypes generation, we also utilized other clustering algorithms, such as the

Table 1. Ablation study on clustering algorithm with BMD enhanced SHOT

Methods/ Datasets	Office-31	Office-Home	VisDA-C
w/o BMD	88.6	71.5	82.9
k-means clustering BMD	**89.4**	**72.5**	85.8
spectral clustering BMD	89.2	72.2	**85.9**

spectral clustering [9] algorithm. The results are summarized in Table. 1. During our experiment, all hyperparameters and architectures are unchanged, though they may be sub-optimal for this variant. From these results, we can see that what is central and effective is to achieve intra-class multiple feature prototype generation, and how this is achieved is not particularly critical. Both the k-means clustering based BMD strategy and the spectral clustering based BMD strategy can yield reasonable and better-than-current results.

B.2 Ablation study on prototype number S

Table 2. Ablation study on prototype number S on VisDA-C

Methods/ S	1	2	3	4	5	6	8
SHOT w/BMD	84.3	85.3	85.8	85.8	85.7	85.7	85.6

To explore the impact of the number S of intra-class multiple feature prototypes on domain adaptation performance, we have conducted an ablation study of S with BMD enhanced SHOT on VisDA-C. The results are summarized in Table. 2. During our experiment, all hyperparameters and architectures are unchanged. In the extreme case of $S = 1$, our intra-class multicentric prototype strategy degenerates to the monocentric strategy. From these results, our intra-class multiple feature prototype strategy is not sensitive to the choice of S, even though $S = 2$, we can obtain much better performance than $S = 1$, which demonstrates the necessity of our intra-class multiple feature prototype strategy for model adaptation.

B.3 Experiments on Generalized Source-free Domain Adaptation

In G-SFDA [10], the authors introduce an novel domain adaptation paradigm called *generalized source-free domain adaptation*, where the learned model need to perform well on both the target and source domains, with only access to current unlabeled target data during model adaptation. To achieve this, they introduce an sparse domain attention mechanism, which produces a binary domain specific attention to activate different feature channels for different domains.

Table 3. Accuracy(%) on VisDA-C validation set with ResNet-101 as backbone under G-SFDA setting. T, S, and H denote accuracy on target, source and generalized domain.

Setting	Methods	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	Avg
T	G-SFDA [10]	95.9	88.1	85.4	72.5	96.1	93.7	88.5	80.6	92.3	92.2	87.6	44.8	84.8
	G-SFDA w/ BMD	95.9	87.5	83.9	75.7	96.5	96.6	91.4	81.8	95.9	88.4	85.1	59.7	**86.5**
S	G-SFDA [10]	99.7	98.7	98.4	80.0	94.6	98.4	76.2	97.8	98.8	99.9	75.6	67.3	90.4
	G-SFDA w/ BMD	99.5	99.7	83.4	67.6	96.1	100.0	78.3	97.6	99.4	98.1	86.5	92.3	**91.5**
H	G-SFDA [10]	97.8	93.1	91.4	76.1	95.3	96.0	81.9	88.4	95.4	95.9	81.2	53.8	87.5
	G-SFDA w/ BMD	97.6	93.2	83.6	71.4	96.3	98.3	84.3	89.0	97.6	93.0	85.8	72.5	**88.9**

To measure the generalized performance, they compute the harmonic mean between source and target accuracy: $H = \frac{2*ACC_S*ACC_T}{ACC_S+ACC_T}$, and ACC_S and ACC_T are respectively the accuracy on source and target test data. In the previous section 5.2, we have conducted experiments to verify the effectiveness of our BMD strategy on G-SFDA in the target domain. In this subsection, we will conduct experiments to demonstrate that our strategy is equally effective for G-SFDA on the source domain. In our experiments, we follow the G-SFDA domain-awareness setup, please refer to the G-SFDA [10] paper for more experimental details. The results on VisDA-C are summarized in Table. 3. From these results, we can conclude that our BMD strategy consistently improves the baseline approach in the source, target, and generalized H-domains. In particular, on the challenging 'truck' class, our BMD strategy can improve the G-SFDA from 44.8% to 59.7% in the target domain, from 67.3% to 92.3% in the source domain, and from 53.8% to 72.5% in the generalized domain.

B.4 Experiments on Multi-Source UDA

Table 4. Multi Source Adaptation on Office-Home(ResNet-50). [$\mathcal{R}$ denotes the rest domains.]

Methods	$\mathcal{R}\rightarrow$Ar	$\mathcal{R}\rightarrow$Cl	$\mathcal{R}\rightarrow$Pr	$\mathcal{R}\rightarrow$Re	Avg
MCD [7]	69.8	59.8	80.9	82.7	73.3
Meta-MCD [3]	70.1	60.5	81.2	83.4	73.8
SImpAI [8]	72.1	62.0	80.3	81.8	74.1
SHOT- ens	73.3	57.7	83.2	83.1	74.3
SHOT w/ BMD -ens	73.5	60.1	84.3	82.7	**75.2**
SHOT -best	73.1	57.0	83.0	82.0	73.8
SHOT w/ BMD -best	73.3	59.5	85.1	82.6	**75.1**
SHOT -worst	68.3	54.1	78.1	80.1	70.2
SHOT w/ BMD -worst	69.6	55.9	77.8	79.9	**70.8**

Table 5. Multi Source Adaptation on Office-31 (ResNet-50). [$\mathcal{R}$ denotes the rest domains.]

Methods	$\mathcal{R}\to$A	$\mathcal{R}\to$D	$\mathcal{R}\to$W	Avg
SHOT -ens	73.8	96.8	93.8	88.1
SHOT w/ BMD -ens	75.0	98.0	94.8	**89.3**
SHOT -best	74.7	99.9	98.4	91.0
SHOT w/ BMD -best	75.6	99.8	97.5	**91.0**
SHOT -worst	74.3	94.0	90.1	86.1
SHOT w/ BMD -worst	75.0	95.6	93.0	**87.9**

Table 6. Multi Target Domain Adaptation on Office-Home

Methods	Ar $\to\mathcal{R}$	Cl$\to\mathcal{R}$	Pr$\to\mathcal{R}$	Re$\to\mathcal{R}$	Avg
CGCT [6]	67.4	68.1	61.6	68.7	66.5
D-CGCT [6]	70.5	71.6	66.0	71.2	69.8
SHOT	68.8	75.3	65.7	67.8	69.4
SHOT w/ BMD	69.3	76.5	66.4	69.5	**70.5**

In addition to the vanilla closed-set domain adaptation, we also verify the effectiveness of our BMD strategy under the multi-source setting. For this setup, we first learn multiple adapted models from different source domain pretrained models. And then, we aggregate the prediction scores of different target models to find the labels with maximum values, we name this method as *-ens* (ensemble). In our comparison, we also compare with the single-source baselines, namely *-best* and *-worst*, which refer to the best adapted source model and the worst model, respectively. We summarize results on Office-Home in Table. 4 and Office-31 in Table. 5. From these results, it is easy to find that our BMD strategy achieves the best results for all settings.

B.5 Experiments on Multi-Target UDA

We also conduct the multi-target domain adaptation experiments to evaluate the effectiveness of our BMD strategy. For this setup, we directly combine these target domains and treat it as the new target domain. The results on Office-Home are summarized in Table. 6. We can see that our BMD strategy can consistently improve SHOT, while also outperforming the existing multi-target domain adaptation approaches.

References

1. Ahmed, S.M., Raychaudhuri, D.S., Paul, S., Oymak, S., Roy-Chowdhury, A.K.: Unsupervised multi-source domain adaptation without access to source data. In: CVPR (2021) 1

2. Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Vaughan, J.W.: A theory of learning from different domains. Machine learning **79**(1), 151–175 (2010) 1
3. Li, D., Hospedales, T.: Online meta-learning for multi-source and semi-supervised domain adaptation. In: ECCV (2020) 3
4. Liang, J., Hu, D., Feng, J.: Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In: ICML (2020) 1
5. Liang, J., Hu, D., Wang, Y., He, R., Feng, J.: Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. IEEE TPAMI (2021) 1
6. Roy, S., Krivosheev, E., Zhong, Z., Sebe, N., Ricci, E.: Curriculum graph co-teaching for multi-target domain adaptation. In: CVPR (2021) 4
7. Saito, K., Watanabe, K., Ushiku, Y., Harada, T.: Maximum classifier discrepancy for unsupervised domain adaptation. In: CVPR (2018) 3
8. Venkat, N., Kundu, J.N., Singh, D.K., Revanur, A., Babu, R.V.: Your classifier can secretly suffice multi-source domain adaptation. In: NeurIPS (2021) 3
9. Von Luxburg, U.: A tutorial on spectral clustering. Statistics and computing **17**(4), 395–416 (2007) 2
10. Yang, S., Wang, Y., van de Weijer, J., Herranz, L., Jui, S.: Generalized source-free domain adaptation. In: ICCV (2021) 2, 3
11. Zou, Y., Yu, Z., Kumar, B., Wang, J.: Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: ECCV (2018) 1
12. Zou, Y., Yu, Z., Liu, X., Kumar, B., Wang, J.: Confidence regularized self-training. In: ICCV (2019) 1