Misalignment-Free Relation Aggregation for Multi-Source-Free Domain Adaptation
Supplementary Material

Hao-Wei Yeh[1], Qier Meng[1], Tatsuya Harada[1,2]
[1]The University of Tokyo [2]RIKEN
yeh@mi.t.u-tokyo.ac.jp, qmeng0517@gmail.com, harada@mi.t.u-tokyo.ac.jp

1. Experiments on 126-class DomainNet Dataset

To validate the proposed method on large-scale dataset, we conduct experiments on the **DomainNet** dataset [4]. As described in the main paper, DomainNet is a large-scale dataset with 345 object classes from 6 domains : clipart images (Cl), infograph images (In), painting images (Pa), quickdraw images (Qu), real-world images (Re), and sketch images (Sk). We follow [5] and select the subset of 126 classes in the 6 domains. Each of the domain takes turns to be the target domain and creates 6 adaptation scenarios in total. ResNet50 [3] is used for the backbone of the source models, and follow the same network architectures and training scheme as the experiments in the main paper. We use the hyper-parameters $\gamma = 1.0$ and $\lambda = 0.1$ for the proposed method. As the results in the main paper, we follow the existing works [1, 2] and report the average accuracy over 3 runs with different random seeds.

The results are summarized in Table 1. As again shown from the results, the proposed method achieves either the highest or second highest accuracy among the existing methods, and the higher average accuracy compared to the existing methods.

2. Examples of Nearest-Neighbor Retrieval with Different Fused Feature Spaces

In Figure 1, we show some of the query-retrieval pairs of the experiment of nearest-neighbor retrieval (Section 5.4 in the main paper). In addition to the higher retrieval rate presented in the main paper, we can also observe that the samples retrieved by "Summing Relations" tend to belong to the ones retrieved by the individual features, which implies better preservation of underlying feature structures of each source model. On the contrary, "Summing Features" tends to retrieve samples that are inconsistent to the ones by individual features, which may due to the loss of underlying structures during fusion.

References

[1] Sk Miraj Ahmed, Dripta S Raychaudhuri, Sujoy Paul, Samet Oymak, and Amit K Roy-Chowdhury. Unsupervised multi-source domain adaptation without access to source data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10103–10112, 2021. 1, 2

[2] Jiahua Dong, Zhen Fang, Anjin Liu, Gan Sun, and Tongliang Liu. Confident anchor-induced multi-source free domain adaptation. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021. 1, 2

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[4] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1406–1415, 2019. 1

[5] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, Trevor Darrell, and Kate Saenko. Semi-supervised domain adaptation via minimax entropy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8050–8058, 2019. 1

Table 1. Accuracies (%) of Object Recognition on 126-class DomainNet Dataset

Method	r→Cl	r→In	r→Pa	r→Qu	r→Re	r→Sk	Average
Source Ensemble	70.3	**30.4**	64.4	14.1	76.1	62.2	52.9
DECISION [1]	67.6	25.8	64.6	13.1	75.8	57.1	50.7
CAiDA [2]	66.3	21.4	63.9	**21.4**	78.1	59.9	51.8
Ours	**78.0**	25.8	**69.5**	15.7	**81.7**	**68.1**	**56.5**

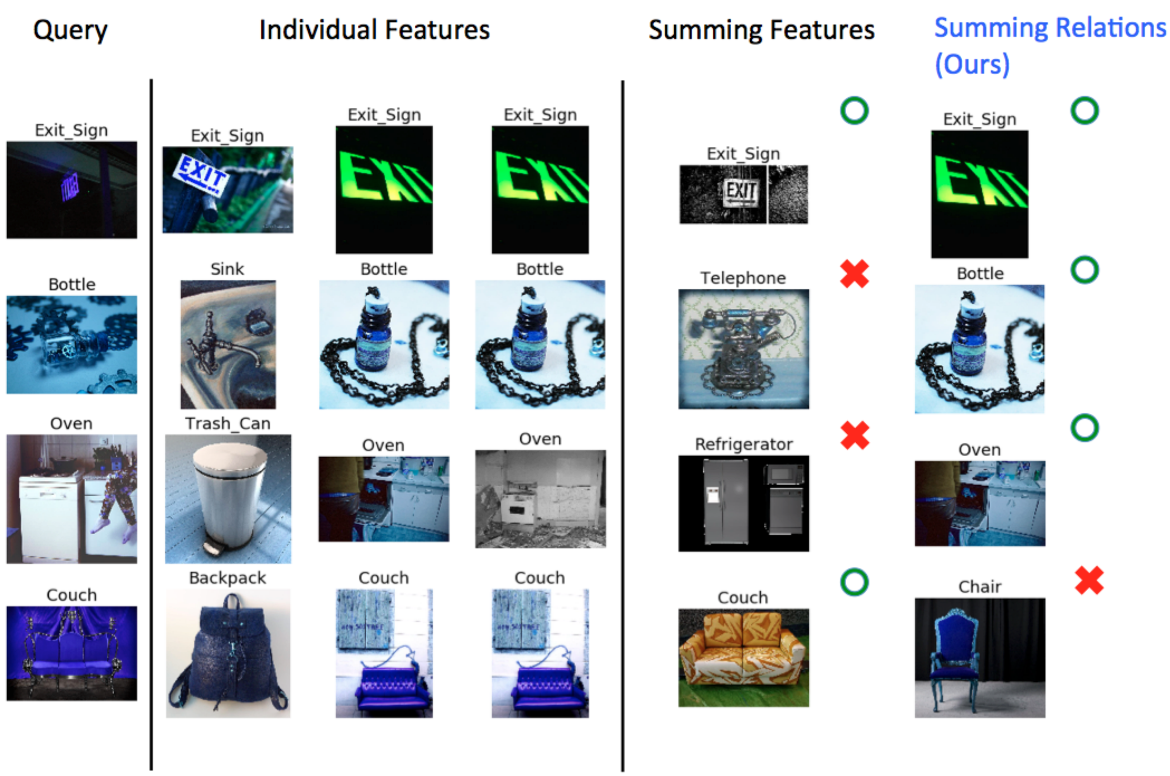

Figure 1. Four examples of the query target samples (first column) from the domain Artworks(A) of OfficeHome, and the corresponding retrieved nearest neighboring target samples in the individual features of trained model from the source domains, Clipart(C), Product(P), and Real-world(R) (each columns of the middle group corresponds to a source model), and the fused feature spaces (last group of columns). The green circle and red cross represent whether the retrieval is a "success" or "failure", respectively. Compared to "Summing Features", "Summing Relations" produce more consistent retrieved samples to the individual features, which implies better preservation of underlying feature structures of each source model.