Learning to Combine: Knowledge Aggregation for Multi-Source Domain Adaptation
Supplementary Material

Hang Wang⋆, Minghao Xu⋆, Bingbing Ni⋆⋆, and Wenjun Zhang

Shanghai Jiao Tong University, Shanghai 200240, China
{wang--hang, xuminghao118, nibingbing, zhangwenjun}@sjtu.edu.cn

1 Detailed Experimental Setups

In this part, we provide detailed experimental setups. For the sake of fair comparison, we follow the backbone setting in [4,2] for different tasks. In our framework, a feature vector is employed to update global prototypes and also serves as query samples, whose dimension varies as the backbone architecture. For different tasks, we list the basic training settings in Table 1.

Table 1: The experimental setups in three different tasks.

dataset	domains	classes	image size	backbone	batch size*	learning rate	training epoch	feature dimension
Digits-five	5	10	32×32	3 conv-2 fc	128	2×10^{-4}	100	2048
Office-31[3]	3	31	252×252	AlexNet	16	5×10^{-5}	100	4096
DomainNet[2]	6	345	224×224	ResNet-101	16	5×10^{-5}	20	2048
PACS[1]	4	7	224×224	ResNet-18	16	5×10^{-5}	100	512

* Batch size here denotes the number of examples sampled from one domain in each iteration.

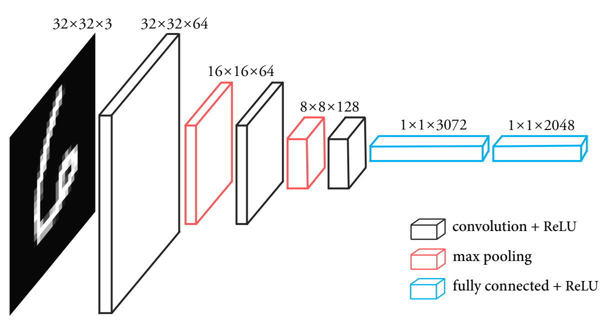

Fig. 1: The network architecture for experiments on Digits-five dataset.

In Figure 1, we provide the detailed network architecture for digits experiments, which mainly follows the design in [2]. Inputted with an image whose spatial size is 32×32, three convolution-based modules produce an $8 \times 8 \times 128$

⋆ Equal contribution.
⋆⋆ Corresponding author: Bingbing Ni.

Table 2: Classification accuracy (%) on *PACS* dataset.

Methods	$\rightarrow$ A	$\rightarrow$ C	$\rightarrow$ S	$\rightarrow$ P	Avg
Source-only	75.97	73.34	64.23	91.65	76.30
MDAN [5]	83.54	82.34	72.42	92.91	82.80
DCTN [4]	84.67	86.72	71.84	95.60	84.71
M^3SDA [2]	84.20	85.68	74.62	94.47	84.74
MDDA [6]	86.73	86.24	77.56	93.89	86.11
LtC-MSDA	**90.19**	**90.47**	**81.53**	**97.23**	**89.85**

feature map. After that, the feature map is flattened, and a 2048-dimensional feature vector is generated by two fully connected layers.

2 Experiments on PACS

Dataset. PACS [1] dataset contains 4 domains, *i.e.* Photo (P), Art paintings (A), Cartoon (C) and Sketch (S). Each domain contains 7 categories, and significant domain shift exists between different domains.

Results. Table 2 reports the performance of our method compared with other works. Source-only denotes the model trained with only source domain data, which serves as the baseline. As shown in the table, the proposed LtC-MSDA model achieves the highest accuracy on all four tasks of PACS dataset, and a 3.74% performance gain is obtained in the term of average accuracy. By combining the knowledges learned from multiple domains, our model show superior performance under huge domain shift settings.

References

1. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Deeper, broader and artier domain generalization. In: IEEE International Conference on Computer Vision. pp. 5542–5550 (2017)
2. Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B.: Moment matching for multi-source domain adaptation. In: IEEE International Conference on Computer Vision (2019)
3. Saenko, K., Kulis, B., Fritz, M., Darrell, T.: Adapting visual category models to new domains. In: European Conference on Computer Vision (2010)
4. Xu, R., Chen, Z., Zuo, W., Yan, J., Lin, L.: Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
5. Zhao, H., Zhang, S., Wu, G., Moura, J.M.F., Costeira, J.P., Gordon, G.J.: Adversarial multiple source domain adaptation. In: Advances in Neural Information Processing Systems (2018)
6. Zhao, S., Wang, G., Zhang, S., Gu, Y., Li, Y., Song, Z.C., Xu, P., Hu, R., Chai, H., Keutzer, K.: Multi-source distilling domain adaptation. In: AAAI Conference on Artificial Intelligence (2020)