

SUPPLEMENTARY MATERIALS FOR SEPREP-NET: MULTI-SOURCE FREE DOMAIN ADAPTATION VIA MODEL SEPARATION AND REPARAMETERIZATION

Anonymous authors

Paper under double-blind review

1 APPENDIX A: DATASET DETAILS

We perform extensive evaluation of SepRep-Net on five benchmark datasets, *i.e.*, **Office-31** Saenko et al. (2010), **Office-Home** Venkateswara et al. (2017), **Digit5** Peng et al. (2019), **Office-Caltech** Gong Boqing & Grauman (2012), and **DomainNet** Peng et al. (2019). **1) Office-31** Saenko et al. (2010) contains 4,652 images from 31 categories collected in the office environment, which forms three domains, *i.e.*, Amazon (**A**), DSLR (**D**) and Webcam (**W**); **2) Office-Home** Venkateswara et al. (2017) consists of 15,500 images, 65 categories from 4 domains, *i.e.*, Art (**Ar**), Clipart (**Cl**), Product (**Pr**) and Real World (**Rw**). The domain gap in Office-Home is much larger than Office-31; **3) Digit5** Peng et al. (2019) is a benchmark dataset for cross-domain digit recognition, with 5 domains, *i.e.*, MNIST (**MT**), USPS (**UP**), SVHN (**SV**), MNIST-M (**MM**) and Synthetic Digits (**SY**); **4) Office-Caltech** Gong Boqing & Grauman (2012) takes the common classes between Caltech-256 and Office-31. It has images from 10 categories and 4 domains, *i.e.*, Amazon (**A**), Caltech (**C**), DSLR (**D**), and Webcam (**W**); **5) DomainNet** Peng et al. (2019) has 345 categories and 6 domains, *i.e.*, Infograph (**I**), Clipart (**C**), Painting (**P**), Quickdraw (**Q**), Real (**R**) and Sketch (**S**).

2 APPENDIX B: IMPLEMENTATION DETAILS

Following previous works, we take ResNet pre-trained on ImageNet as the feature extractor to recognize objects in the real world. For digit recognition, we adopt the network structure in Liang et al. (2020) and train it from scratch. We use the same bottleneck layer and task-specific classifier as Liang et al. (2020).

We take one domain as the target domain and view the remaining domains as source domains, which forms multiple tasks in each dataset. For real-world objects, we follow the conventions of image size and augmentation in Ahmed et al. (2021); Liang et al. (2020). For digit datasets, we rescale images to 32×32 and convert gray-scale images to RGB. We obtain source models following the training scheme in Ahmed et al. (2021); Liang et al. (2020) to ensure a fair comparison.

The target model is trained with the SGD optimizer. The training schedule follows the settings in Liang et al. (2020). Moreover, the learning rate of the bottleneck and classifier is $10\times$ larger than the feature extractor. The whole framework is trained end-to-end. We implement our experiments with Pytorch, run each experiment 5 times and report the average results.

3 APPENDIX C: EXPERIMENT RESULTS

3.1 OFFICE-CALTECH

Besides Office-31 Saenko et al. (2010), Office-Home Venkateswara et al. (2017), Digit5 Peng et al. (2019), and DomainNet Peng et al. (2019) that are reported in the main paper, on Office-Caltech Gong Boqing & Grauman (2012), as shown in Table 1, our method can also be readily integrated into DECISION Ahmed et al. (2021), CAiDA Dong et al. (2021), and SHOT Liang et al. (2020). By reassembling three source models into one target model, our method consistently achieves better performance than the vanilla method with much less computational costs. Compared

Table 1: Source Accuracy (S) (%), Target Accuracy (T) (%), H-Score (H) and Model Efficiency on **Office-Caltech** dataset. ResNet-50 is adopted in experiments. A, C, D, W indicate different domains (A: Amazon, C: Caltech, D: DSLR, W: Webcam). SHOT-ens indicates the performance of model ensemble with all models that are adapted via SHOT. KD indicates knowledge distillation.

METHOD	FLOPS	C,D,W \rightarrow A			A,D,W \rightarrow C			A,C,W \rightarrow D			A,C,D \rightarrow W			Avg		
		S	T	H	S	T	H	S	T	H	S	T	H	S	T	H
ResNet He et al. (2016)	12.3	-	88.7	-	-	85.4	-	-	98.2	-	-	99.1	-	-	92.9	-
DAN Long et al. (2015)	12.3	-	91.6	-	-	89.2	-	-	99.1	-	-	99.5	-	-	94.8	-
DCTN Xu et al. (2018)	12.3	-	92.7	-	-	90.2	-	-	99.0	-	-	99.4	-	-	95.3	-
MCD Saito et al. (2018)	12.3	-	92.1	-	-	91.5	-	-	99.1	-	-	99.5	-	-	95.6	-
M3SDA Peng et al. (2019)	12.3	-	94.5	-	-	92.2	-	-	99.2	-	-	99.5	-	-	96.4	-
DECISION Ahmed et al. (2021)	12.3	93.1	95.9	94.5	93.1	95.9	94.5	88.1	100.0	93.7	91.4	99.6	95.3	91.4	98.0	94.6
DECISION + KD Hinton et al. (2015)	4.1	88.3	96.0	92.0	90.7	95.7	93.1	87.9	99.4	93.3	85.0	99.7	91.8	88.0	97.7	92.5
DECISION + SepRep-Net	4.1	94.8	96.1	95.4	95.1	96.1	95.6	90.2	100.0	95.6	92.5	99.8	96.0	93.2	98.0	95.5
CAiDA Liang et al. (2020)	12.3	96.0	96.8	96.4	96.2	97.1	96.6	98.0	100.0	99.0	98.1	99.8	98.9	97.1	98.4	97.7
CAiDA + KD Hinton et al. (2015)	4.1	91.5	96.6	94.0	92.3	96.8	94.5	94.6	100.0	97.2	95.9	99.8	97.8	93.6	98.3	95.9
CAiDA + SepRep-Net	4.1	97.2	97.1	97.1	96.8	97.1	96.9	98.4	100.0	99.2	97.9	99.7	98.8	97.6	98.5	98.0
SHOT-ens Liang et al. (2020)	12.3	98.7	95.7	97.2	98.0	95.8	96.9	98.3	96.8	97.5	98.1	99.6	98.8	98.3	97.0	97.6
SHOT-ens + KD Hinton et al. (2015)	4.1	93.4	95.6	94.5	95.0	95.7	95.3	94.2	96.6	95.4	90.3	96.5	93.3	93.2	96.1	94.6
SHOT + SepRep-Net	4.1	99.3	95.9	97.6	98.1	95.6	96.8	97.7	97.5	97.6	98.1	99.7	98.9	98.3	97.2	97.7

with knowledge distillation (KD), our method also shows better performance, especially in source domains, which indicates stronger generability.

REFERENCES

- 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 *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
- Jiahua Dong, Zhen Fang, Anjin Liu, Gan Sun, and Tongliang Liu. Confident anchor-induced multi-source free domain adaptation. In *Advances in Neural Information Processing Systems*, volume 34, 2021.
- Sha Fei Gong Boqing, Shi Yuan and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.
- Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International Conference on Machine Learning*, pp. 6028–6039, 2020.
- Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I. Jordan. Learning transferable features with deep adaptation networks. In *International Conference on Machine Learning*, 2015.
- Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *IEEE International Conference on Computer Vision*, pp. 1406–1415, 2019.
- Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *European Conference on Computer Vision*, 2010.
- Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep Hashing Network for Unsupervised Domain Adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

Ruijia Xu, Ziliang Chen, Wangmeng Zuo, Junjie Yan, and Liang Lin. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.