Supplement Material of "Meta-causal Learning for Single Domain Generalization"

Jin Chen[1*], Zhi Gao[1*], Xinxiao Wu[1,2†], Jiebo Luo[3]
[1]Beijing Key Laboratory of Intelligent Information Technology,
School of Computer Science & Technology, Beijing Institute of Technology, China
[2]Guangdong Laboratory of Machine Perception and Intelligent Computing,
Shenzhen MSU-BIT University, China
[3]Department of Computer Science, University of Rochester, Rochester NY 14627, USA
{chen_jin,gaozhi_2017,wuxinxiao}@bit.edu.cn, jluo@cs.rochester.edu

1. Future work

In this work, the factor set is defined manually, which may not capture all causes of the domain shift. In future work, we are going to discover latent factors so as to analyze the domain shift more flexibly and comprehensively.

2. Ablation Study of Eq.8

We add an ablation study by removing Eq.8 from our method ("Ours w/o Eq.8"). From the results shown in Table 1, "Ours w/o Eq.8" performs worse than "Ours". The reason is that Eq.8 restricts each feature mapping to address a specific domain shift, facilitating the reduction of combined domain shift in Eq.7.

Table 1. Ablation studies (%) of Eq.8 on PACS with ResNet-18.

Method	Artpaint	Cartoon	Sketch	Photo	Avg
Ours w/o Eq.8	74.37	75.65	59.71	55.10	66.21
Ours	**77.13**	**80.14**	**62.55**	**59.60**	**69.86**

3. Network Architecture.

For the Digits dataset, we use ConvNet [9] as the backbone [11, 17]. Each feature mapping is built with two FC layers (512→1024) with ReLU following the first FC layer. For the CIFAR-10 dataset, we use WRN [22] with 16 layers and widen factor 4 as the backbone [11, 17]. Each feature mapping is built with four FC layers (512→512→512→256) with ReLU following the first three FC layer. For the PACS dataset, we use ResNet-18 [7] pretrained on ImageNet as the backbone [6, 20]. Each feature mapping is built with four FC layers (1024→1024→1024→2048) with ReLU following the first three fc layer. For all tasks, the effect-to-weight network is built with two FC layers ($|\mathcal{Y}|\rightarrow 10\times|\mathcal{Y}|\rightarrow 1$) with ReLU after the first FC layer, where $|\mathcal{Y}|$ is the number of categories.

* Jin Chen and Zhi Gao are co-first authors.
† Corresponding author: Xinxiao Wu.

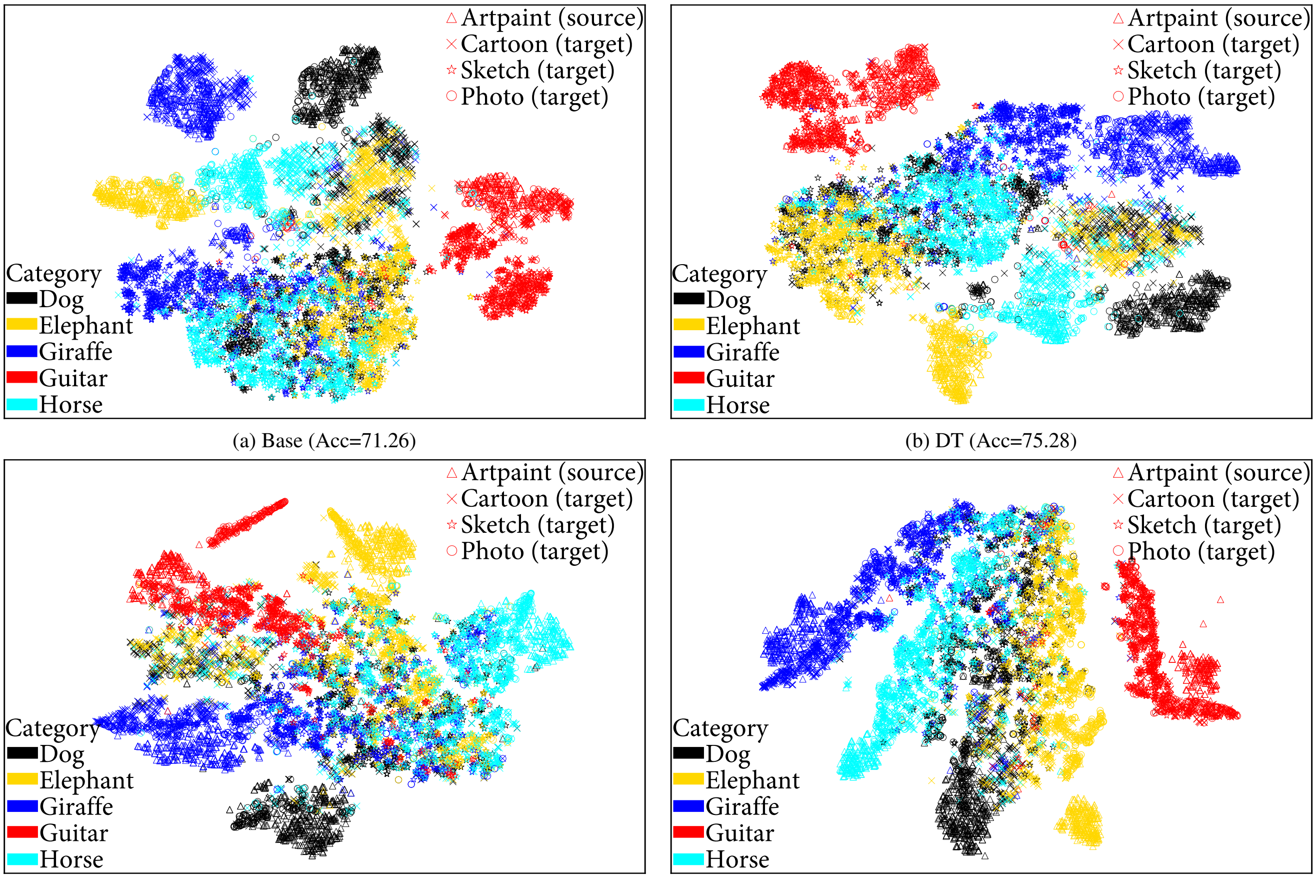
Table 2. Leave-one-domain-out results (%) on PACS with ResNet-50. One domain (name in column) is used as the target domain and the other three domains are used as source domains.

Method	Artpaint	Cartoon	Photo	Sketch	Avg
MetaReg [1]	87.20	79.20	97.60	70.30	83.60
MASF [5]	82.89	80.49	95.01	72.29	82.67
EISNet [19]	86.64	81.53	97.11	78.07	85.84
RSC [8]	87.89	82.16	97.92	83.35	87.83
FACT [21]	89.63	81.77	96.75	84.46	88.15
MatchDG [14]	85.61	82.12	**97.94**	78.76	86.11
CSG-ind [12]	88.60	84.60	97.80	81.10	88.03
CIRL [13]	**90.67**	84.30	97.84	87.68	90.12
Ours	90.58	**85.11**	97.60	**88.62**	**90.48**

4. Results on Multiple Domain Generalization

To further evaluate the effectiveness of our method, we also conduct experiments of multi-source domain generalization with ResNet50 as backbone on PACS. We employ the leave-one-domain-out protocol following existing multi-source domain generalization [13, 21]. We compare our method with most related methods that introduces causal inference into generalization (MatchDG [14], CSG-ind [12], CIRL [13]), and existing popular domain generalization methods (MetaReg [1], GUD [18], Epi-FCR [10], MASF [5], JiGen [2], DMG [3], DDAIG [24], CSD [16], L2A-OT [25], EISNet [19], RSC [8], ME-ADA [23], MMLD [15], L2D [20], FACT [21]).

Table 2 shows the leave-one-domain-out results on the PACS dataset with ResNet-50 as backbone. From the re-



(a) Base (Acc=71.26)

(b) DT (Acc=75.28)

(c) DTA (Acc=71.64)

(d) Ours (Acc=77.13)

Figure 1. Feature visualization on the PACS dataset with ResNet-18 as backbone. "△", " × ", "☆", and " ○ " denote the features of the source domain (Artpaint) and three target domains (Cartoon, Sketch, Photo), respectively. Different colors denote different categories as shown in the legend.

sults, it is noteworthy that our method achieves the state-of-the-art results in terms of the overall metric "Avg" although we not utilize the domain labels during training. In particular, when Cartoon or Sketch are used as the target domain, the domain shift is larger than other tasks due to the totally different styles of Cartoon and the highly abstracted shapes of Sketch. In these challenging tasks, our method not only beats all popular domain generalization methods but also surpasses the methods of introducing causal inference (MatchDG [14], CSG-ind [12], CIRL [13]), clearly showing the advantages of analyzing the causes of the domain shift by causal inference.

5. Feature Visualization

In Figure 1, we visualize the data distributions of the learned features by "Base", "DT", "DTA" and our method on the PACS dataset with Artpaint as the source domain, where ResNet-18 is used as backbone. For clarity, we only show the first five categories from both the source and unseen target domains using t-SNE embeddings [4]. It is noteworthy to observe that comparing with this three variants of our method, our method aligns the source features with the target features of unseen target domains well, further verifying the superiority of the new *simulate-analyze-reduce* paradigm.

References

[1] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. Metareg: Towards domain generalization using meta-regularization. volume 31, pages 1006–1016, 2018. 1

[2] Fabio M Carlucci, Antonio D'Innocente, Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. Domain generalization by solving jigsaw puzzles. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2229–2238, 2019. 1

[3] Prithvijit Chattopadhyay, Yogesh Balaji, and Judy Hoffman. Learning to balance specificity and invariance for in and out of domain generalization. In *European Conference on Computer Vision (ECCV)*, pages 301–318. Springer, 2020. 1

[4] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *International Conference on Machine Learning (ICML)*, pages 647–655, 2014. 2

[5] Qi Dou, Daniel Coelho de Castro, Konstantinos Kamnitsas, and Ben Glocker. Domain generalization via model-agnostic learning of semantic features. volume 32, pages 6447–6458, 2019. 1

[6] Xinjie Fan, Qifei Wang, Junjie Ke, Feng Yang, Boqing Gong, and Mingyuan Zhou. Adversarially adaptive normalization for single domain generalization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8208–8217, 2021. 1

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 1

[8] Zeyi Huang, Haohan Wang, Eric P Xing, and Dong Huang. Self-challenging improves cross-domain generalization. In *European Conference on Computer Vision (ECCV)*, pages 124–140. Springer, 2020. 1

[9] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989. 1

[10] Da Li, Jianshu Zhang, Yongxin Yang, Cong Liu, Yi-Zhe Song, and Timothy M Hospedales. Episodic training for domain generalization. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1446–1455, 2019. 1

[11] Lei Li, Ke Gao, Juan Cao, Ziyao Huang, Yepeng Weng, Xiaoyue Mi, Zhengze Yu, Xiaoya Li, and Boyang Xia. Progressive domain expansion network for single domain generalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 224–233, 2021. 1

[12] Chang Liu, Xinwei Sun, Jindong Wang, Haoyue Tang, Tao Li, Tao Qin, Wei Chen, and Tie-Yan Liu. Learning causal semantic representation for out-of-distribution prediction. volume 34, pages 6155–6170, 2021. 1, 2

[13] Fangrui Lv, Jian Liang, Shuang Li, Bin Zang, Chi Harold Liu, Ziteng Wang, and Di Liu. Causality inspired representation learning for domain generalization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8046–8056, 2022. 1, 2

[14] Divyat Mahajan, Shruti Tople, and Amit Sharma. Domain generalization using causal matching. In *International Conference on Machine Learning (ICML)*, pages 7313–7324. PMLR, 2021. 1, 2

[15] Toshihiko Matsuura and Tatsuya Harada. Domain generalization using a mixture of multiple latent domains. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 34, pages 11749–11756, 2020. 1

[16] Vihari Piratla, Praneeth Netrapalli, and Sunita Sarawagi. Efficient domain generalization via common-specific low-rank decomposition. In *International Conference on Machine Learning (ICML)*, pages 7728–7738. PMLR, 2020. 1

[17] Fengchun Qiao, Long Zhao, and Xi Peng. Learning to learn single domain generalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12556–12565, 2020. 1

[18] Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. volume 31, pages 5339–5349, 2018. 1

[19] Shujun Wang, Lequan Yu, Caizi Li, Chi-Wing Fu, and Pheng-Ann Heng. Learning from extrinsic and intrinsic supervisions for domain generalization. In *European Conference on Computer Vision (ECCV)*, pages 159–176. Springer, 2020. 1

[20] Zijian Wang, Yadan Luo, Ruihong Qiu, Zi Huang, and Mahsa Baktashmotlagh. Learning to diversify for single domain generalization. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 834–843, 2021. 1

[21] Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. A fourier-based framework for domain generalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14383–14392, 2021. 1

[22] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *British Machine Vision Conference (BMVC)*. British Machine Vision Association, 2016. 1

[23] Long Zhao, Ting Liu, Xi Peng, and Dimitris Metaxas. Maximum-entropy adversarial data augmentation for improved generalization and robustness. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 14435–14447, 2020. 1

[24] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Deep domain-adversarial image generation for domain generalisation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 34, pages 13025–13032, 2020. 1

[25] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Learning to generate novel domains for domain generalization. In *European Conference on Computer Vision (ECCV)*, pages 561–578. Springer, 2020. 1