

# DGMamba: Domain Generalization via Generalized State Space Model – Supplementary Material

Anonymous Authors

## 1 EXPERIMENT DETAILS

**Dataset Details.** In this section, we provide details of the 5 commonly used DG benchmarks in Table 1. The data in these datasets originates from various sources with distinct characteristics, such as hand-drawn illustrations, software-composited images, object-centered photographs, and scene-centered shots.

**Evaluation Protocols.** Following the standard evaluation protocols [3, 5, 12, 22, 25], we report the generalization performance based on the train-domain validation, *i.e.*, selecting one domain as the target domain and training on the remaining domains. Each source domain is split with an 80%/20% train/validation ratio. The validation parts from these source domains collectively constitute the validation set used for the model evaluation.

**Hyperparameters for DomainNet.** The benchmark DomainNet [14] presents a great challenge, containing 586575 images. Training 10,000 iterations represents less than two complete epochs on DomainNet, which is insufficient for the model to converge effectively. Consequently, recent state-of-the-art DG works [3, 9, 23] increase the iterations to 15000 with a batch size of 32 for each domain. To make a fair comparison while considering the constraints posed by GPU, our proposed DGMamba undergoes 80000 iterations with a batch size of 6 for each source domain. The corresponding initial learning rate is searched in [7e-4, 8.5e-4].

**Overall Architecture.** For an input image  $x$ , it undergoes initial processing into patches  $x_i \in x$  utilizing a convolution neural network with layer normalization. These patches are subsequently scanned to be processed by four Mamba blocks. Down-sampling is applied after the first three Mamba blocks to generate feature maps with different resolutions. Finally, prediction is performed using a linear classifier subsequent to the layer normalization, average pooling, and flattening operations.

## 2 ADDITIONAL EXPERIMENTS

**Comparison on More Complex dataset.** To further assess the efficacy of mitigating distribution shifts in large-scale benchmarks, we test the proposed DGMamba on DomainNet [14] and compare it with state-of-the-art DG methods. As indicated in Table 1, DomainNet comprises 586575 images categorized into 345 classes from six domains. As illustrated by Table 2, our proposed DGMamba demonstrates a substantial improvement of 2.7% compared to the state-of-the-art approach in terms of average generalization performance across diverse domains. Remarkably, our DGMamba attains the best performance in five out of the six domains. These findings underscore the superiority of the proposed DGMamba in tackling distribution shifts in real-world applications.

**Comparison across Five Benchmarks.** In Table 3, we present a summary of the generalization performance results across five DG benchmarks with training-validation selection. Notably, our proposed DGMamba remarkably outperforms the state-of-the-art DG approaches across all benchmarks, yielding an average gain

Table 1: Statistics of DG benchmarks.

Dataset	Domain	# image	# image	# class
PACS [10]	Art	2048	9991	7
	Photo	1670		
	Clipart	2344		
	Sketch	3929		
VLCS [4]	Caltech	1415	10729	5
	LabelMe	2656		
	SUN	3282		
	PASCAL	3376		
OfficeHome [18]	Art	2427	15588	65
	Clipart	4365		
	Product	4439		
	Real	4357		
TerraIncognita [1]	L100	4741	24330	10
	L38	9736		
	L43	3970		
	L46	5883		
DomainNet [14]	Clipart	48129	586575	345
	Infograph	51605		
	Painting	72266		
	Quickdraw	172500		
	Real	172947		
	Sketch	69128		

of 4.4%. The significant enhancements across diverse scenarios from different benchmarks underscore the excellent ability of our DGMamba to tackle distribution shifts.

**Visualization for Distribution Gaps.** To visually demonstrate that our proposed DGMamba could effectively address the distribution shifts between diverse domains, we employ the t-SNE technique [17] to examine the representation’s distribution gaps across domains. We conduct experiments on PACS with ‘Art’ as the target domain. Figure 1 presents the visualization results based on the CNN-based method iDAG [7], ViT-based method GMoE [9], VMamba [11], and our proposed DGMamba, respectively. Notably, our DGMamba demonstrates a reduced distribution gap between the source and target domains compared to these state-of-the-art methods, indicating its superiority in learning domain-invariant features. In contrast, the representation generated by iDAG exhibits

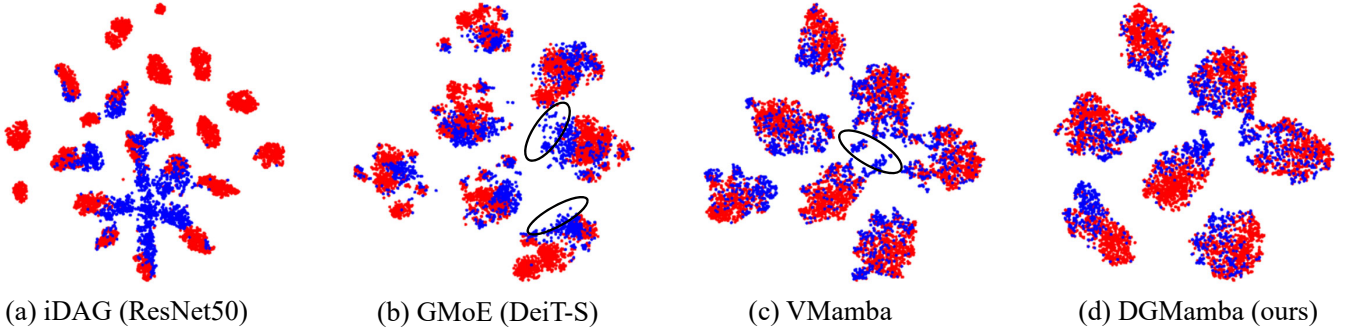
**Table 2: Results on DomainNet with our DGMamba.**

Method	Backbone	Params.	Clipart	Infograph	painting	Quickdraw	Real	Sketch	Avg.(↑)
VREx [8](ICML’2021)	ResNet50	23M	47.3	16.0	35.8	10.9	49.6	42.0	33.6
MTL [2] (JMLR’2021)	ResNet50	23M	57.9	18.5	46.0	12.5	59.5	49.2	40.6
Mixstyle [24] (ICLR’2021)	ResNet50	23M	51.9	13.3	37.0	12.3	46.1	43.4	34.0
SagNet [13](CVPR’2021)	ResNet50	23M	57.7	19.0	45.3	12.7	58.1	48.8	40.3
ARM [21] (NeurIPS’2021)	ResNet50	23M	49.7	16.3	40.9	9.4	53.4	43.5	35.5
SWAD [3] (NeurIPS’2021)	ResNet50	23M	66.0	22.4	53.5	16.1	65.8	55.5	46.5
PCL [20] (CVPR’2022)	ResNet50	23M	67.9	24.3	55.3	15.7	66.6	56.4	47.7
SAGM [19] (CVPR’2023)	ResNet50	23M	64.9	21.1	51.5	14.8	64.1	53.6	45.0
iDAG [7] (ICCV’2023)	ResNet50	23M	67.9	24.2	55.0	16.4	66.1	56.9	47.7
GMDG [16] (CVPR’2024)	ResNet50	23M	63.4	22.4	51.4	13.4	64.4	52.4	44.6
SDViT [15] (ACCV’2022)	DeiT-S	22M	63.4	22.9	53.7	15.0	67.4	52.6	45.8
GMoE [9] (ICLR’2023)	DeiT-S	34M	<b>68.2</b>	24.7	55.7	16.3	69.1	55.4	48.3
DGMamba (ours)	VMamba-T	22M	67.0	<b>27.9</b>	<b>56.5</b>	<b>18.4</b>	<b>69.5</b>	<b>57.9</b>	<b>49.6</b>

**Table 3: Comparison of state-of-the-art DG methods with our DGMamba. Out-of-domain generalization performance on five commonly used benchmarks is reported. The best results are bolded.**

Method	Backbone	Params.	Dataset					Avg.(↑)
			PACS	VLCS	OfficeHome	TerraIncognita	DomainNet	
VREx [8] (ICML’2021)	ResNet50	23M	84.9	78.3	66.4	46.4	33.6	61.9
RSC [6] (ECCV’2020)	ResNet50	23M	85.2	77.1	65.5	46.6	38.9	62.7
MTL [2] (JMLR’2021)	ResNet50	23M	84.6	77.2	66.4	45.6	40.6	62.9
Mixstyle [24] (ICLR’2021)	ResNet50	23M	85.2	77.9	60.4	44.0	34.0	60.3
SagNet [13] (CVPR’2021)	ResNet50	23M	86.3	77.8	68.1	48.6	40.3	64.2
ARM [21] (NeurIPS’2021)	ResNet50	23M	85.1	77.6	64.8	45.5	35.5	61.7
SWAD [3] (NeurIPS’2021)	ResNet50	23M	88.1	79.1	70.6	50.0	46.5	66.9
PCL [20] (CVPR’2022)	ResNet50	23M	88.7	78.0	71.6	52.1	47.7	67.6
SAGM [19] (CVPR’2023)	ResNet50	23M	86.6	80.0	70.1	48.8	45.0	66.1
iDAG [7] (ICCV’2023)	ResNet50	23M	88.8	76.9	71.8	46.1	47.7	66.3
GMDG [16] (CVPR’2024)	ResNet50	23M	85.6	79.2	70.7	50.1	44.6	66.0
SDViT [15] (ACCV’2022)	DeiT-S	22M	86.3	78.9	71.5	44.3	45.8	65.4
GMoE [9] (ICLR’2023)	DeiT-S	34M	86.7	78.0	72.4	45.6	48.3	66.2
DGMamba (ours)	VMamba-T	22M	<b>91.2</b>	<b>80.8</b>	<b>77.0</b>	<b>54.6</b>	<b>49.6</b>	<b>70.6</b>

● source domain      ● target domain

**Figure 1: Visualizations with t-SNE embeddings [17] illustrating features’ distribution gaps between the source and target domains generated by (a) iDAG [7], (b) GMoE [9], (c) VMamba [11], and (d) DGMamba (ours), respectively. Our proposed DGMamba displays the superior feature alignment.**

**Table 4: Comparison of state-of-the-art DG methods with our DGMamba with diverse backbones. The best results are bolded.**

Method	Backbone	Params.	Art	Cartoon	Photo	Sketch	Avg.(↑)
iDAG [7] (ICCV’2023)	ResNet50	23M	90.8	83.7	98.0	82.7	88.8
iDAG [7] (ICCV’2023)	ResNet101	41M	89.0	84.9	98.3	84.7	89.2
GMoE [9] (ICLR’2023)	DeiT-S	34M	89.4	83.9	99.1	74.5	86.7
GMoE [9] (ICLR’2023)	DeiT-B	133M	91.0	84.0	99.3	82.7	89.2
DGMamba (ours)	VMamba-T	22M	91.3	87.0	99.0	87.3	91.2
DGMamba (ours)	VMamba-S	47M	94.1	87.8	99.6	<b>89.0</b>	92.6
DGMamba (ours)	VMamba-B	83M	<b>95.1</b>	<b>89.2</b>	<b>99.8</b>	87.9	93.0

a noticeable distinction between the source and target domains, indicating its weakness in tackling distribution shifts. For GMoE, there exist target features that are away from the source features, as indicated by the black circles. Besides, the representations produced by GMoE exhibit poor intra-class compactness. For VMamba, the ability to align features is inferior to our DGMamba, manifesting by the distant target features marked by the black circle, and the distinction between classes is not as clear as that in DGMamba.

**Performance with Diverse Backbones.** To fully unleash the potential of our proposed DGMamba, we investigate the impact of utilizing larger backbones, *i.e.*, stacking more Mamba layers or increasing the embedding dimension to facilitate capturing genuine features. Specifically, we conduct experiments on PACS utilizing VMamba-S and VMamba-B. VMamba-S comprises 4 blocks, each including 2, 2, 15, and 2 Mamba layers, with an embedding dimension of 96. VMamba-B maintains the same Mamba blocks and layers as VMamba-S, with an increased embedding dimension of 128. The generalization performances are concluded in Table 4, demonstrating that deeper Mamba architectures or larger embedding dimensions could enhance the model generalizability. Furthermore, our proposed DGMamba demonstrates superior generalization performance compared to CNN-based or ViT-based models, while maintaining comparable or fewer parameters. These results underscore the effectiveness of DGMamba in mitigating domain shifts.

## REFERENCES

- [1] Sara Beery, Grant Van Horn, and Pietro Perona. 2018. Recognition in terra incognita. In *Proceedings of the European Conference on Computer Vision*. 456–473.
- [2] Gilles Blanchard, Aniket Anand Deshmukh, Ürün Dogan, Gyemin Lee, and Clayton Scott. 2021. Domain generalization by marginal transfer learning. *The Journal of Machine Learning Research* 22, 1 (2021), 46–100.
- [3] Junbum Cha, Sanghyuk Chun, Kyungjae Lee, Han-Cheol Cho, Seunghyun Park, Yunsung Lee, and Sungrae Park. 2021. Swad: Domain generalization by seeking flat minima. *Advances in Neural Information Processing Systems* 34 (2021), 22405–22418.
- [4] Chen Fang, Ye Xu, and Daniel N Rockmore. 2013. Unbiased metric learning: On the utilization of multiple datasets and web images for softening bias. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1657–1664.
- [5] Ishaan Gulrajani and David Lopez-Paz. 2020. In search of lost domain generalization. In *International Conference on Learning Representations*.
- [6] Zeyi Huang, Haohan Wang, Eric P Xing, and Dong Huang. 2020. Self-challenging improves cross-domain generalization. In *Proceedings of the European Conference on Computer Vision*. Springer, 124–140.
- [7] Zenan Huang, Haobo Wang, Junbo Zhao, and Nenggan Zheng. 2023. iDAG: Invariant DAG searching for domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 19169–19179.
- [8] David Krueger, Ethan Caballero, Joern-Henrik Jacobsen, Amy Zhang, Jonathan Binias, Dinghuai Zhang, Remi Le Priol, and Aaron Courville. 2021. Out-of-distribution generalization via risk extrapolation (rex). In *International Conference on Machine Learning*. PMLR, 5815–5826.
- [9] Bo Li, Yifei Shen, Jingkang Yang, Yezhen Wang, Jiawei Ren, Tong Che, Jun Zhang, and Ziwei Liu. 2023. Sparse mixture-of-experts are domain generalizable learners. In *International Conference on Learning Representations*.
- [10] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M. Hospedales. 2017. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5543–5551.
- [11] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, and Yunfan Liu. 2024. Vmamba: Visual state space model. *arXiv preprint arXiv:2401.10166* (2024).
- [12] Shaocong Long, Qianyu Zhou, Chenhao Ying, Lizhuang Ma, and Yuan Luo. 2023. Rethinking domain generalization: Discriminability and generalizability. *arXiv preprint arXiv:2309.16483* (2023).
- [13] Hyeonseob Nam, Hyunjae Lee, Jongchan Park, Wonjun Yoon, and Donggeun Yoo. 2021. Reducing domain gap by reducing style bias. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8690–8699.
- [14] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. 2019. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1406–1415.
- [15] Maryam Sultana, Muzammal Naseer, Muhammad Haris Khan, Salman Khan, and Fahad Shahbaz Khan. 2022. Self-distilled vision transformer for domain generalization. In *Proceedings of the Asian Conference on Computer Vision*. 3068–3085.
- [16] Zhaorui Tan, Xi Yang, and Kaizhu Huang. 2024. Rethinking multi-domain generalization with a general learning objective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- [17] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 11 (2008).
- [18] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. 2017. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5385–5394.
- [19] Pengfei Wang, Zhaoxiang Zhang, Zhen Lei, and Lei Zhang. 2023. Sharpness-aware gradient matching for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3769–3778.
- [20] Xufeng Yao, Yang Bai, Xinyun Zhang, Yuechen Zhang, Qi Sun, Ran Chen, Ruiyu Li, and Bei Yu. 2022. PCL: Proxy-based contrastive learning for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7097–7107.
- [21] Marvin Zhang, Henrik Marklund, Nikita Dhawan, Abhishek Gupta, Sergey Levine, and Chelsea Finn. 2021. Adaptive risk minimization: Learning to adapt to domain shift. *Advances in Neural Information Processing Systems* 34 (2021), 23664–23678.
- [22] Yifan Zhang, Xue Wang, Kexin Jin, Kun Yuan, Zhang Zhang, Liang Wang, Rong Jin, and Tieniu Tan. 2023. AdaNPC: Exploring non-parametric classifier for test-time adaptation. In *International Conference on Machine Learning*. PMLR, 41647–41676.
- [23] YiFan Zhang, xue wang, Jian Liang, Zhang Zhang, Liang Wang, Rong Jin, and Tieniu Tan. 2023. Free lunch for domain adversarial training: Environment label smoothing. In *International Conference on Learning Representations*.
- [24] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. 2021. Domain generalization with mixstyle. In *International Conference on Learning Representations*.
- [25] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. 2024. Mixstyle neural networks for domain generalization and adaptation. *International Journal of Computer Vision* 132, 3 (2024), 822–836.