Supplementary Materials for DandelionNet: Domain Composition with Instance Adaptive Classification for Domain Generalization

Lanqing Hu[1,2], Meina Kan[1,2], Shiguang Shan[1,2,3], Xilin Chen[1,2]
[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100090, China
[2]University of Chinese Academy of Sciences, Beijing, 100090, China
[3]Peng Cheng Laboratory, Shenzhen, 518055, China
lanqing.hu@vipl.ict.ac.cn,{kanmeina,sgshan,xlchen}@ict.ac.cn

1. Supplementary Experiments and Analyses

In this supplementary section, more analyses on how to achieve better results with different hyperparameter r, and feature/deviation visualization are given. Additionally, full results on domain generalization benchmarks are also supplemented.

1.1. More Analyses on Tuning Deviation Norm Hyperparameter r

In our proposed DandelionNet, to reduce the difficulty of optimization and also avoid trial solution, the deviation norm is simply constrained to the same for all domains and classes. But in an oracle model, the norm values are probably different. In this section, we simply conduct trials about setting different r for varied domain on $L38, L43, L46 \rightarrow L100$ experiment in *TerraIncognita* benchmark. The results are shown in Table 1. As we firstly observe that the best accuracy is achieved in this experiment when $r = 0.1$ and $r = 0.05$, here r is automatically searched in this scope. It is observed that after finely tuning r for each source domain, the accuracy on unseen target domain can be further improved, e.g., about 2.0% in this trial experiment, which indicates that specifying degrees of deviation for different domains is reasonable, and *our work would be quite promising with more flexible domain specific and semantic category specific classifier weight deviation*.

1.2. More Visualization Results on Different Benchmarks

We also visualize feature space and learnt deviation in VLCS which contains smaller number of classes (5), and OfficeHome which contains larger number of classes (65) in Figure 1. As illustrated, features from the same class are clustered together as well as with individual variations. Besides, the automatically optimized deviation also shows domain composition in these multiple benchmarks, demonstrating that feature and our deviation are optimized as expected.

Table 1: More flexible hyperparameter r on $L38, L43, L46 \rightarrow L100$ in TerraIncognita. The light blue are the best and second best results with the same r, and the yellow one is the best result after simply tuning different r for each source domain.

Deviation norm set for sources			Acc
L38	L43	L46	L100
0.1	0.1	0.1	63.1
0.1	0.1	0.05	63.1
0.1	0.05	0.1	67.7
0.1	0.05	0.05	65.3
0.05	0.1	0.1	65.7
0.05	0.1	0.05	64.1
0.05	0.05	0.1	61.9
0.05	0.05	0.05	65.6

1.3. Full Results on Domain Generalization Benchmarks

In this section, full results with the best hyperparameter $r = 0.1$ on the five domain generalization benchmarks are shown in Tables 2, 3, 4, 5, and 6 for more detailed verification of this proposed method.

1.4. Results with different model architectures

To demonstrate the versatility of our DandelionNet for multiple model architectures besides ResNet50, supplementary experiments with smaller ResNet18 and larger transformer-based ViT-B/16 are shown in Table 7.

1.5. More validatations on SFDA under less intra-domain variations

For source free domain adaptation (SFDA), In most experiment and reality settings, the source and target domains are with varied variation. But here we also add exemplifications on the setting where the intra-domain variations are similar, i.e., the most difference between domains are the

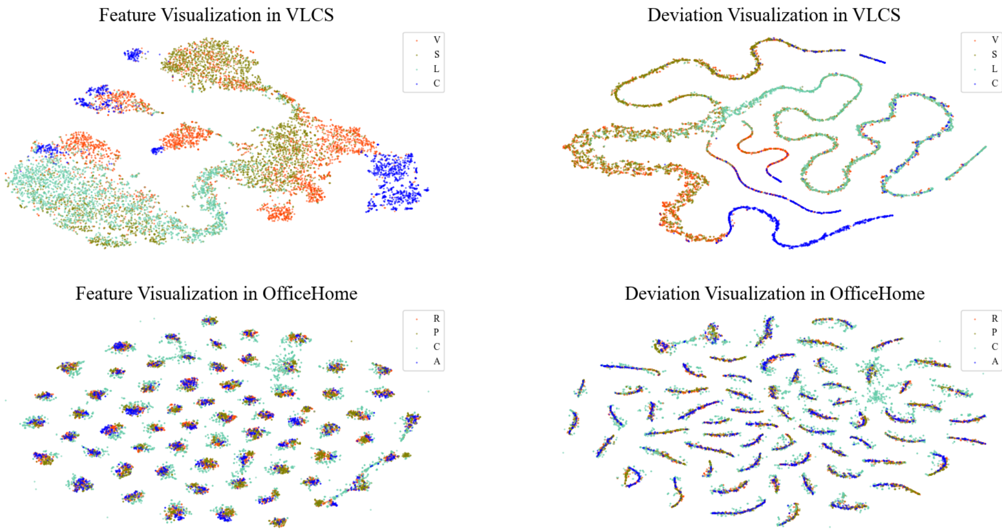

Figure 1: Feature and deviation visualization on VLCS ($L, S, V \rightarrow C$) and OfficeHome ($C, P, R \rightarrow A$), where the blue stand for target domain.

Table 2: Out-of-domain accuracies ($\%$) on PACS.

Method	A	C	P	S	Avg
CDANN [12]	84.6	75.5	96.8	73.5	82.6
IRM[1]	84.8	76.4	96.7	76.1	83.5
ERM [14]	85.7	77.1	97.4	76.6	84.2
MASF[8]	82.9	80.5	95.0	72.3	82.7
MetaReg [3]	87.2	79.2	97.6	70.3	83.6
GroupDRO [13]	83.5	79.1	96.7	78.3	84.4
VREx[7]	86.0	79.1	96.9	77.7	84.9
MLDG [11]	85.5	80.1	97.4	76.6	84.9
ARM[17]	86.8	76.8	97.4	79.3	85.1
RSC[9]	85.4	79.7	97.6	78.2	85.2
Mixstyle[18]	86.8	79.0	96.6	78.5	85.2
SWAD[4]	89.3	83.4	97.3	82.5	88.1
EoA[2]	90.5	83.4	98.0	82.5	88.6
Ours	87.8	86.5	96.8	85.8	89.2

Table 3: Out-of-domain accuracies ($\%$) on VLCS.

Method	C	L	S	V	Avg
CDANN [12]	97.1	65.1	70.7	77.1	77.5
IRM[1]	98.6	64.9	73.4	77.3	78.6
ERM [14]	98.0	64.7	71.4	75.2	77.3
I-Mixup [16]	98.3	64.8	72.1	74.3	77.4
GroupDRO [13]	97.3	63.4	69.5	76.7	76.7
VREx[7]	98.4	64.4	74.1	76.2	78.3
MLDG[11]	97.4	65.2	71.0	75.3	77.2
ARM[17]	98.7	63.6	71.3	76.7	77.6
Mixstyle[18]	98.6	64.5	72.6	75.7	77.9
SWAD [4]	98.8	63.3	75.3	79.2	79.1
EoA[2]	99.1	63.1	75.9	78.3	79.1
Ours	99.1	70.2	77.2	80.0	81.6

main domain shift, The results in Table 8 indicate that our method can achieve good performance in not only the more complicated scenarios where both intra- and inter- domain variations are significant but also the simpler scenarios where only the domain shift is dominant.

References

[1] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019. 2, 3

[2] D. Arpit, H. Wang, Y. Zhou, and C. Xiong. Ensemble of averages: Improving model selection and boosting performance in domain generalization. In *NeurIPS*, 2022. 2, 3

[3] Y. Balaji, S. Sankaranarayanan, and R. Chellappa. Metareg: Towards domain generalization using meta-regularization. In *NeurIPS*, 2018. 2, 3

[4] J. Cha, S. Chun, Kyungjae Lee, and et al. Swad: Domain generalization by seeking flat minima. In *NeurIPS*, 2021. 2, 3

[5] J. Cha, K. Lee, S. Park, and S. Chun. Domain generalization by mutual-information regularization with pre-trained models. In *ECCV*, 2022. 3

Table 4: Out-of-domain accuracies (%) on OfficeHome.

Method	A	C	P	R	Avg
CDANN [12]	61.5	50.4	74.4	76.6	65.7
IRM[1]	58.9	52.2	72.1	74.0	64.3
ERM [14]	63.1	51.9	77.2	78.1	67.6
I-Mixup [16]	62.4	54.8	76.9	78.3	68.1
GroupDRO [13]	60.4	52.7	75.0	76.0	66.0
VREx[7]	60.7	53.0	75.3	76.6	66.4
MLDG[11]	61.5	53.2	75.0	77.5	66.8
ARM[17]	58.9	51.0	74.1	75.2	64.8
Mixstyle[18]	51.1	53.2	68.2	69.2	60.4
SWAD [4]	66.1	57.7	78.4	80.2	70.6
EoA[2]	69.1	59.8	79.5	81.5	72.5
Ours	65.8	58.6	78.0	79.7	70.5

Table 5: Out-of-domain accuracies (%) on TerraIncognita.

Method	L100	L38	L43	L46	Avg
CDANN [12]	47.0	41.3	54.9	39.8	45.8
IRM[1]	54.6	39.8	56.2	39.6	47.6
ERM [14]	54.3	42.5	55.6	38.8	47.8
I-Mixup [16]	59.6	42.2	55.9	33.9	47.9
GroupDRO [13]	41.2	38.6	56.7	36.4	43.2
VREx[7]	48.2	41.7	56.8	38.7	46.4
MLDG[11]	54.2	44.3	55.6	36.9	47.8
ARM[17]	49.3	38.3	55.8	38.7	45.5
Mixstyle[18]	54.3	34.1	55.9	31.7	44.0
SWAD [4]	55.4	44.9	59.7	39.9	50.0
EoA[2]	57.8	46.5	61.3	43.5	52.3
Ours	63.1	48.7	60.0	46.4	54.5

Table 6: Out-of-domain accuracies (%) on DomainNet.

Method	clp	inf	pnt	qdr	rel	skt	Avg
CDANN [12]	54.6	17.3	43.7	12.1	56.2	45.9	38.3
IRM[1]	48.5	15.0	38.3	10.9	48.2	42.3	33.9
ERM [14]	63.0	21.2	50.1	13.9	63.7	52.0	44.0
I-Mixup [16]	55.7	18.5	44.3	12.5	55.8	48.2	39.2
GroupDRO [13]	47.2	17.5	33.8	9.3	51.6	40.1	33.3
VREx[7]	47.3	16.0	35.8	10.9	49.6	42.0	33.6
MLDG[11]	59.1	19.1	45.8	13.4	59.6	50.2	41.2
ARM[17]	49.7	16.3	40.9	9.4	53.4	43.5	35.5
Mixstyle[18]	51.9	13.3	37.0	12.3	46.1	43.4	34.0
MetaReg [3]	59.8	25.6	50.2	11.5	64.6	50.1	43.6
DMG [6]	65.2	22.2	50.0	15.7	59.6	49.0	43.6
SWAD [4]	66.0	22.4	53.5	16.1	65.8	55.5	46.5
EoA[2]	65.9	23.4	55.3	16.5	66.4	57.1	47.4
Ours	66.5	24.6	55.8	16.6	67.8	57.5	48.1

Table 7: Out-of-domain accuracies (%) on DomainBed with ResNet18 and ViT-B/16 .

ResNet18	PACS	VLCS	OfficeHome	TerraIncognita	
XDED[10]	83.8	74.8	65.0	42.5	
Ours	85.3	78.2	66.5	48.2	
ViT-B/16	PACS	VLCS	OfficeHome	TerraIncognita	DomainNet
MIRO [5]	95.6	82.2	82.5	54.3	54.0
Ours	96.3	83.4	85.8	57.0	57.1

Table 8: Source free domain adaptation on CIFAR10C at level 5 of corruption (Err %).

WRN-28	gn	sn	in	db	gb	mb	zb	snow
TENT[15]	24.8	23.5	33.0	12.0	31.8	13.7	10.8	15.9
Ours	21.1	19.6	27.2	12.2	29.4	13.9	11.8	16.2
WRN-28	fros	fog	brt	contr	et	pixel	jpg	
TENT[15]	16.2	13.7	7.9	12.1	22.0	17.3	24.2	
Ours	15.3	12.5	8.5	10.2	22.3	16.1	20.1	

[6] P. Chattopadhyay, Y. Balaji, and J. Hoffman. Learning to balance specificity and invariance for in and out of domain generalization. In *ECCV*, 2020. 3

[7] J.-H. Jacobsen D. Krueger, E. Caballero and et al. Out-of-distribution generalization via risk extrapolation. In *ICML*, 2021. 2, 3

[8] Q. Dou, D. C. Castro, K. Kamnitsas, and B. Glocker. Domain generalization via model-agnostic learning of semantic features. In *NeurIPS*, 2019. 2

[9] Z. Huang, H. Wang, E. P Xing, and D. Huang. Self-challenging improves cross-domain generalization. In *ECCV*, 2020. 2

[10] K. Lee, S. Kim, and S. Kwak. Cross-domain ensemble distillation for domain generalization. In *ECCV*, 2022. 3

[11] D. Li, Y. Yang, Y. Song, and T. Hospedales. Learning to generalize: Meta-learning for domain generalization. In *AAAI*, 2018. 2, 3

[12] Y. Li, M. Gong, X. Tian, and et al. Domain generalization via conditional invariant representations. In *AAAI*, 2018. 2, 3

[13] S. Sagawa, P. W. Koh, and P. Liang T. B. Hashimoto. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. In *ICLR*, 2020. 2, 3

[14] R. Vedantam, D. Lopez-Paz, and D. J. Schwab. An empirical investigation of domain generalization with empirical risk minimizers. In *NeurIPS*, 2021. 2, 3

[15] D. Wang, E. Shelhamer, S. Liu, and et al. Tent: Fully test-time adaptation by entropy minimization. In *ICLR*, 2021. 3

[16] M. Xu, J. Zhang, B. Ni, and et al. Adversarial domain adaptation with domain mixup. In *AAAI*, 2020. 2, 3

[17] M. Zhang, H. Marklund, A. Gupta, S. Levine, and C. Finn. Adaptive risk minimization: A meta-learning approach for tackling group shift. *arXiv preprint arXiv:2007.02931*, 2020. 2, 3

[18] K. Zhou, Y. Yang, Y. Qiao, and T. Xiang. Domain generalization with mixstyle. In *ICLR*, 2021. 2, 3