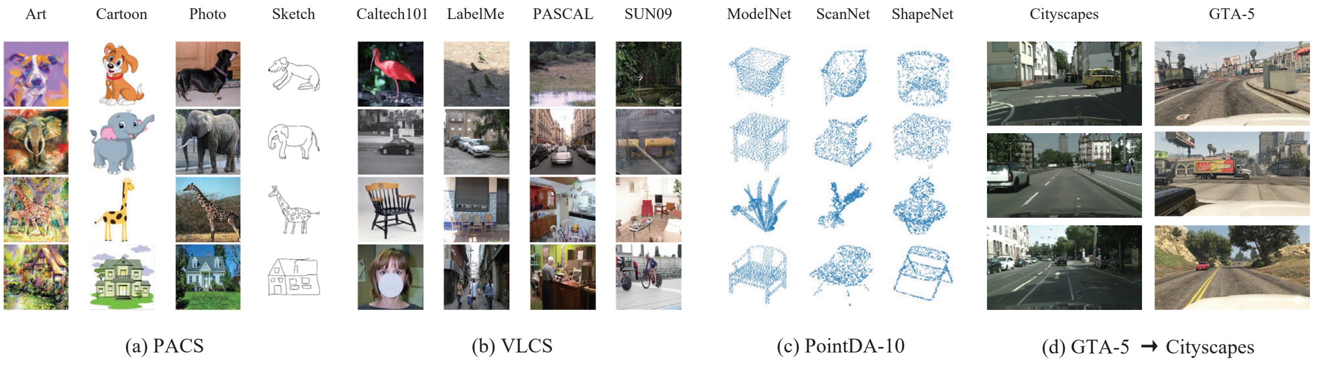
Modality-Agnostic Debiasing for Single Domain Generalization — Supplementary Material

Sanqing Qu, Yingwei Pan, Guang Chen*, Ting Yao, Changjun Jiang, Tao Mei

Tongji University, HiDream.ai Inc.

{2011444, guangchen, cjjiang}@tongji.edu.cn, {panyw.ustc, tingyao.ustc}@gmail.com, tmei@hidream.ai



(a) PACS (b) VLCS (c) PointDA-10 (d) GTA-5 → Cityscapes

Figure 1. Examples from four domain generalization benchmarks that manifest different types of domain shifts. In (a), image styles differences are the main source for domain shifts. In (b), the domain shifts mainly correspond to the changes of environments and viewpoints. In (c), the domain shifts are solely derived from the geometric differences. In (d), driving scenes changes are the main reason for domain shifts.

1. More Details about Datasets

In the main paper, we have validated the effectiveness of our Modality-Agnostic Debiasing (MAD) framework in a variety of single domain generalization (single-DG) scenarios with different modalities, including recognition on 1D texts, 2D images, 3D point clouds, and semantic segmentation on 2D images. Here we provide more details about the adopted datasets in the main paper. The statistics are listed in Table 1.

In an effort to qualitatively show the domain shifts in different benchmarks, we further illustrate some examples in Figure 1. One major observation is that the domain shifts vary a lot between benchmarks. For example, the domain shifts in images (Figure 1 (a), (b), (d)) mostly result from the changes for image contexts, styles, and viewpoints. In point clouds (Figure 1 (c)), the domain shifts primary correspond to geometric variations. Existing single-DG methods are commonly designed for images by devising various data augmentation algorithms to introduce various textures and image styles, making them modality-specific and only applicable to the single modality inputs of images. In contrast, MAD proposes to directly enhance the classifier's ability to identify domain-specific features while emphasizing the learning of domain-generalized features. In this way, a versatile modality-agnostic single-DG paradigm is established by completely eliminating the need for modality-specific data augmentations. MAD is also appealing due to the fact that it can be seamlessly incorporated into existing single-DG methods to further boost up performances.

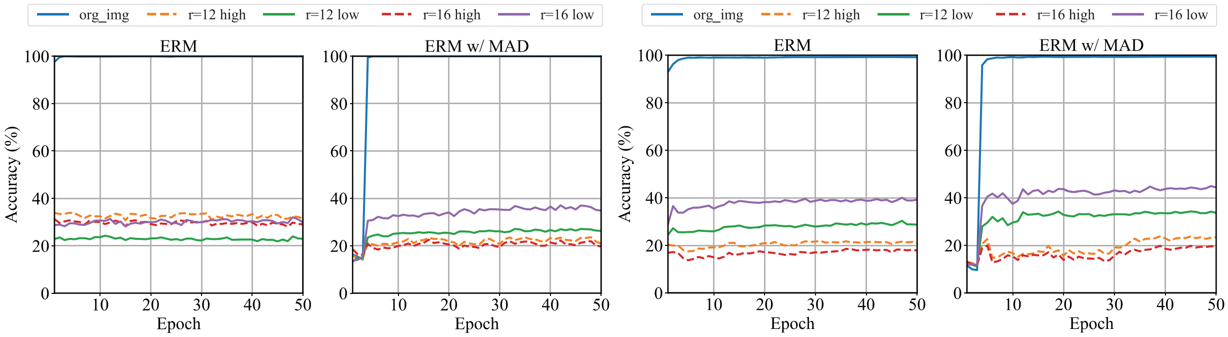
2. More Results for Low-Frequency Component vs. High-Frequency Component

For images, Low-frequency component (LFC) is commonly considered as domain-generalized features, while High-frequency component (HFC) is regarded as domain-specific features [8]. Here, we provide more results to support the capacity of MAD enforcing classifiers to pay more attention to domain-generalized features, i.e., LFC. Here we conduct additional experiments in the "Photo" and "Art" domains on PACS benchmark. Implementation details are the same as in the main paper. That is, for each instance in the validation subset, we decompose the image into LFC and HFC w.r.t different radius threshold r via applying Fourier transform and inverse Fourier transform. Then, we train the ERM and the ERM w/ MAD, separately, and evaluate them on LFC and HFC. The results are summarized in

[*] Corresponding author

Table 1. The statistics of benchmark datasets.

Dataset	#Domain	#Class	#Sample	Description	Reference
PACS	4	7	9,991	Art, Cartoon, Photos, Sketches.	[3]
VLCS	4	5	10,729	Caltech101, LabelMe, SUN09, VOC2007.	[7]
PointDA	3	10	32,788	ShapeNet, ScanNet, ModelNet.	[5]
AmazonReview	4	2	8,000	DVDs, Kitchen, Electronics, Books.	[1]
GTA5→ Cityscapes	2	19	29,966	Semantic segmentation generalization from synthetic images to realistic images.	[2, 6]



(a) LFC vs. HFC in "Photo" domain on PACS

(b) LFC vs. HFC in "Art" domain on PACS

Figure 2. Comparisons of ERM and ERM w/ MAD training curves on low-frequency component (LFC) and high-frequency component (HFC). Experiments are conducted on PACS. All curves in this figure are from validation samples.

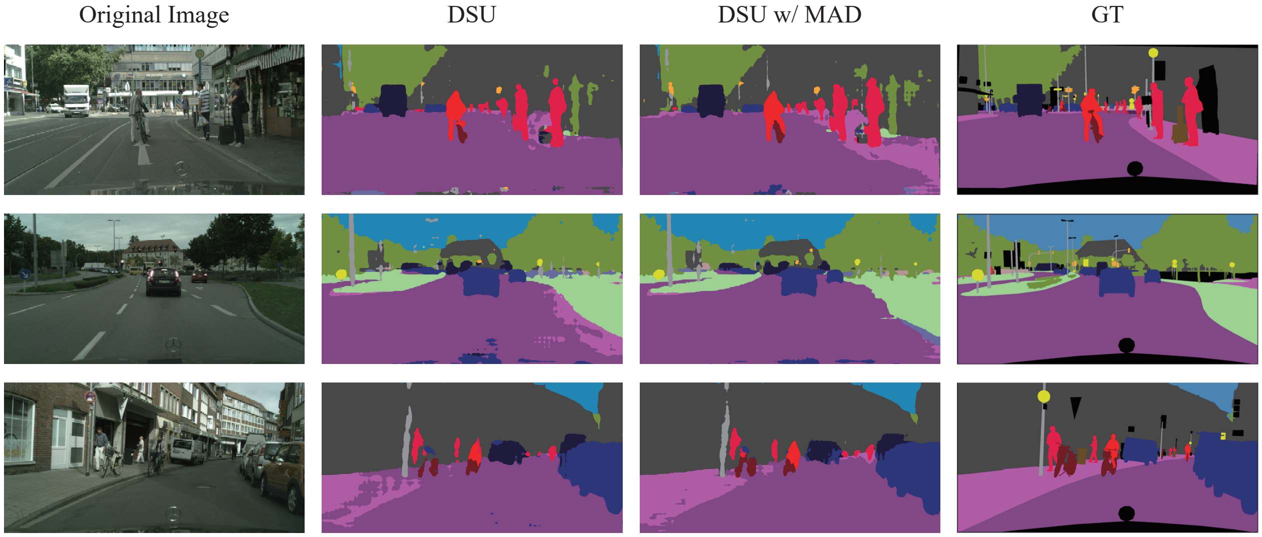


Figure 3. Semantic segmentation illustration on unseen domain Cityscapes with model trained on GTA-5.

Figure 2, where $r = 12/16low$ represents the LFC and $r = 12/16high$ depicts the HFC. As shown in this figure, we can conclude that MAD consistently encourages the classifier focus more on those domain-generalized features.

3. More Results for Semantic Segmentation Visualization

Semantic segmentation models often suffer from performance degradation due to scenario changes. We exhibit more visualization results of semantic segmentation in Figure 3. These examples further demonstrate the effectiveness of MAD when integrated into existing data-augmentation based methods (e.g., DSU [4]).

References

[1] John Blitzer, Ryan McDonald, and Fernando Pereira. Domain adaptation with structural correspondence learning. In *EMNLP*, 2006. 2

[2] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, 2016. 2

[3] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In *ICCV*, 2017. 2

[4] Xiaotong Li, Yongxing Dai, Yixiao Ge, Jun Liu, Ying Shan, and Ling-Yu Duan. Uncertainty modeling for out-of-distribution generalization. In *ICLR*, 2022. 2

[5] Can Qin, Haoxuan You, Lichen Wang, C-C Jay Kuo, and Yun Fu. Pointdan: A multi-scale 3d domain adaption network for point cloud representation. In *NeurIPS*, 2019. 2

[6] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *ECCV*, 2016. 2

[7] Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *CVPR*, 2011. 2

[8] Haohan Wang, Xindi Wu, Zeyi Huang, and Eric P Xing. High-frequency component helps explain the generalization of convolutional neural networks. In *CVPR*, 2020. 1