FSDR: Frequency Space Domain Randomization for Domain Generalization (Supplemental Materials)

A. The definition of band-pass filter $\mathcal{B}_p()$

To define the Average Decomposition Band-Pass Filter $\mathcal{B}_p(\cdot)$, we first consider a gray-scale (one channel) image $x \subset \mathbb{R}^{n\times n}$. Then we use $x(i,j)$ to index the value of x at position (i,j), and we use (c_1, c_2) to denote the centroid (*i.e.*, $c1 = c2 = n/2$). The band-pass filtering consists of two filtering steps, namely, low-reject filtering and high-reject filtering that are defined as follows:

$$x' = x,$$
$$x' = \begin{cases} 0, & \text{if } d((i,j),(c_1,c_2)) < r^{low}\frac{n}{Q}, \\ x'(i,j), & \text{otherwise}, \end{cases}$$
$$x' = \begin{cases} 0, & \text{if } d((i,j),(c_1,c_2)) > r^{up}\frac{n}{Q}, \\ x'(i,j), & \text{otherwise}, \end{cases}$$

where $d(\cdot,\cdot)$ denotes Euclidean distance; r^{low}/r^{up} denotes the lower/upper threshold; Q denotes how many components (*i.e.*, 64) the input is supposed to be decomposed into.

We conduct the Average Decomposition Band-Pass Filter $\mathcal{B}_p(\cdot)$ by implementing above functions 64 times with 64 lower/upper thresholds $(r^{low}, r^{up}) \in \{(0,1),(1,2),(2,3)......(63,64)\}$, and stacking them to get $x' \subset \mathbb{R}^{n\times n\times 64}$. For the last band-pass filtering, we do not perform high-reject filtering to assign all highest information to the last components, as the filter is a circle while image is a square.

For rectangle input image, we first up-sample the length of short side to that of long side before this process and down-sample it back to the original length after processing. If x has more than one channel (*i.e.*, RGB input image $x \subset \mathbb{R}^{n\times n\times 3}$), the procedure operates on every channel of pixels independently (*i.e.*, output $x' \subset \mathbb{R}^{n\times n\times 3\times 64}$).

B. Object detection training detail

We follow DA [1] and Faster RCNN [2] to conduct the domain generalizable object detection experiments. We set the shorter side of all images to 500 and train the networks with learning rate 0.001 for 50K iterations, then with learning rate 0.0001 for 20K more iterations and reported the final performance. All models are trained with this scheduling and we reported the performance trained after 70K iterations. A momentum of 0.9 and a weight decay of 0.0005 is used in the experiments.

C. Image classification details

We implement spectrum analysis over a synthetic-to-real domain generalizable image classification task (*i.e.*, cropped SYNTHIA to ImageNet). We chose this task with three considerations: (1) It avoids using any target domain (*i.e.*, Cityscapes, Mapillary and BDD) annotations (segmentation and detection); (2) ImageNet is a large-scale dataset that can represent the real world distribution, on which the learning should be more generic and robust. We use ResNet-101 as the backbone and run the experiments with 100 epochs with the ADAM optimizer with learning rate t 10^{-4}. The experiments are conducted over the five common classes (*i.e.*, person, car, bus, motorbike, bike) between SYNTHIA and ImageNet.

References

[1] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3339–3348, 2018. 1

[2] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015. 1