

Supplementary Materials for Frequency Decomposition to Tap the Potential of Single Domain for Generalization

A. Example of Frequency Slices

Fig. 1 shows the original images and the frequency slices of PACS. It can be seen in the first column that the style of different domains is very different. And different column shows different frequency bands. The low-frequency band has more colors and energy and the high-frequency band has lines and less energy except for Sketch.

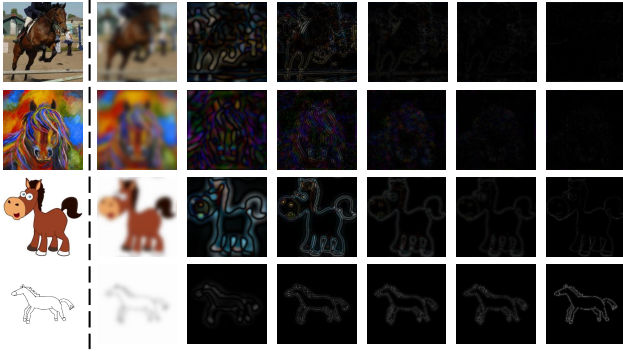


Figure 1. PACS image frequency slices of six frequency bands. The four rows are photo, art painting, cartoon, and sketch. The seven columns are original images and F1~F6 frequency slices.

B. Accuracy of Different Frequency Slices

Table 1 is the accuracy value of three training settings consistent with Fig. 3 in the main paper. From the value, we can see the performance gap of the three methods in each frequency band. The accuracy of the model trained on the original image is 28.89% ~ 56.84% lower than that trained on each frequency slice, which is a very big gap. It indicates that training on the original image can not learn the information of each frequency component well, while there is effective information in every frequency band indeed. The accuracy of our method is very close to that of training on each frequency band, which shows that our method has learned effective information in each frequency band well.

Table 1. Accuracy(%) of different training methods with testing on the frequency slices. The first two lines are trained on the original photo samples and on each frequency component of photo images. The third line is the method proposed by us. Each column means a frequency component for test with F1 lowest and F6 highest.

	F1	F2	F3	F4	F5	F6
Original	43.86	50.88	68.19	39.41	23.28	30.64
Filtered	93.57	89.47	97.08	86.55	80.12	84.21
Ours	84.21	87.72	95.91	88.3	79.53	80.12

C. Similar Information Between Frequency Slices

Table 2. Accuracy(%) of photo domain slices. They are trained with the left column and tested with the top row. The best accuracy of each row is in **bold faces**.

Photo	F1	F2	F3	F4	F5	F6
F1	91.22	29.82	30.99	23.39	19.3	18.13
F2	30.41	88.89	78.36	35.67	19.88	26.32
F3	14.04	54.39	95.91	72.51	26.32	40.94
F4	23.98	35.67	87.13	88.89	75.44	79.53
F5	21.05	21.05	45.03	77.78	83.63	78.36
F6	19.88	16.37	32.75	73.1	79.53	83.63

Table 3. Accuracy(%) of sketch domain slices. They are trained with the left column and tested with the top row. The best accuracy of each row is in **bold faces**.

Sketch	F1	F2	F3	F4	F5	F6
F1	90.7	50.75	56.28	57.29	27.64	31.66
F2	26.38	92.71	85.43	85.18	52.51	54.78
F3	24.37	83.17	96.23	89.95	63.32	74.37
F4	34.17	82.16	85.23	95.23	76.88	77.64
F5	19.10	57.54	85.68	94.72	94.72	81.91
F6	19.10	71.11	82.16	87.19	86.43	93.22

Different frequency bands may contain similar information, which we called cross-information. The proportion

of cross information between different frequency bands of different domains may be different. For example, Photo has bright colors and borders, while Sketch only has object outlines without colors. We tested in these two extreme cases. Specifically, we train the model in each frequency band and test it with all frequency bands. The higher accuracy means the information contained in the two frequency bands is more similar, and more cross-information.

Based on the results shown in Tab. 2 and Tab. 3, there is more similar information in closer frequency bands. The accuracy of photo between frequency bands is almost below 80%, while many of them in the sketch domain is more than 80%. There is less cross-information between various frequency bands of Photo than Sketch. This is maybe a reason why the performance of training on Photo is better than on Sketch. With less cross-information, the model can learn more information with less disturbance in each frequency band and achieve better performance.

D. Visualization of Frequency Band 3 Features

We use t-SNE to visualize the distribution of the F3 slices features of the photo domain, associate with column F3 in Tab. 1. The models of (a) and (b) are trained on the original photo samples and on frequency slices of F3. The model of (c) is the method proposed by us. Figure 2 shows the result. It can be seen that the ERM method cannot distinguish various categories well, while training on the F3 and our method can extract features with good classification. This is consistent with the accuracy in Tab. 1.

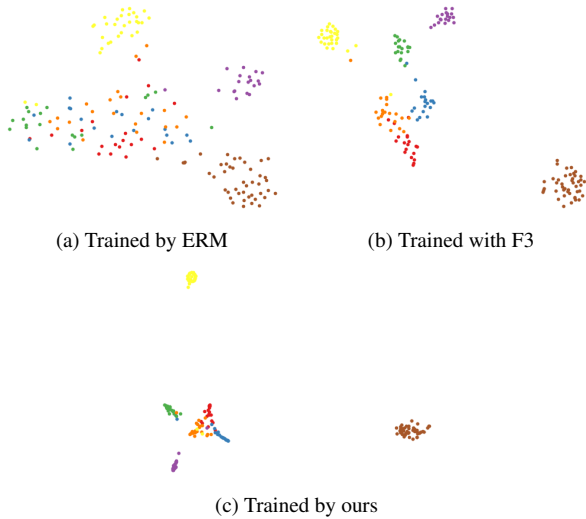


Figure 2. The t-SNE visualizations of Photo F3 frequency slices feature distribution, associated with column F3 in Tab. 1. (a) The model is trained on the original photo images. (b) The model is trained on F3 frequency slices. (c) The model is trained on Photo with our approach. Features with the same semantic label are drawn in the same color.

E. Broader Impact

In this paper, we provide a solution for single domain generalization, which enables the model to achieve a better generalization effect with a single domain training set. It improves the adaptability of the model and reduces the cost of collecting multi-source data. Meanwhile, compared with the method of generating new domain images, our approach of mining the information of the data itself eliminates the possibility of introducing unrealistic information or offensive content. According to our knowledge, our work may not adversely affect the moral aspects and future social consequences.