

Frequency Decomposition to Tap the Potential of Single Domain for Generalization

BMVC 2024 Submission # 740

1 Extra Experiments

1.1 Example of Frequency Slices

Figure 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.

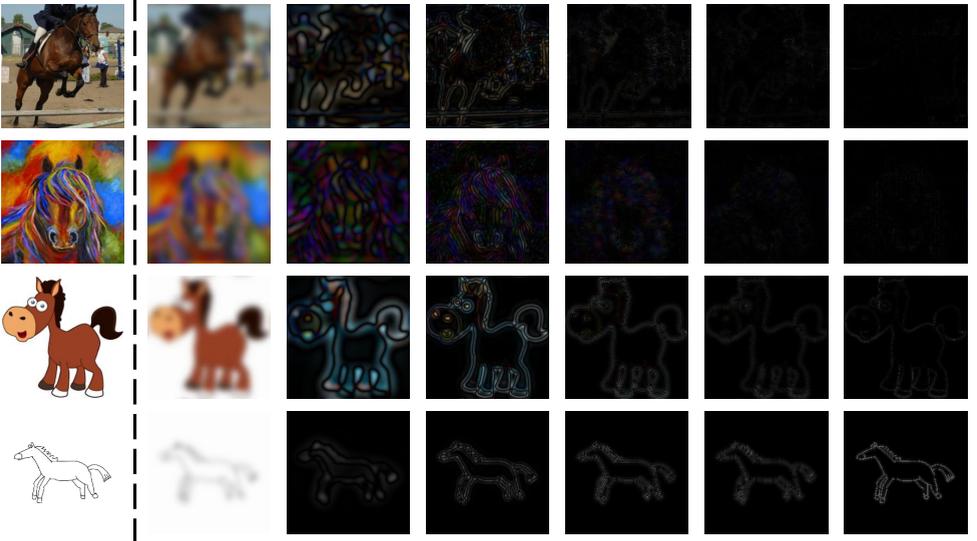


Figure 1: PACS image frequency slices of six frequency bands. The four rows are photo, art painting, cartoon, and sketch. The seven columns are original images and F1~F6 frequency slices.

Method	A	C	S	Avg.
pass branch	56.87	51.11	68.35	58.42
stop branch	58.72	41.60	55.37	51.89
two same branches	66.36	29.39	33.22	42.99
ours w/o L_{cons}	64.16	41.64	58.34	54.71
ours	66.41	53.07	74.10	64.52

Table 1: Accuracy(%) of models trained on single branch or without L_{cons} . Photo is the source domain, and A, C, and S are the target domains.

1.2 Ablations

The proposed framework contains two branches, and the similarity between their outputs is calculated as a loss to assist in learning the effective information in the two frequency domains. To verify the effectiveness of each part of the framework, we conducted ablation experiments, and the results can be found in Table 1.2.

About two branches. There is no structural difference between the two branches. If the branch is taken out separately, the difference between the two is that they receive complementary frequency slices as input. Both of the single-branch experiment outperforms ERM. We also tested two-branch ResNet18 as a control experiment. And the results are also better than ERM but not comparable to ours.

An interesting observation is that the accuracy of the pass branch is higher than the stop branch. The only difference between the two single branches is that the samples in the pass branch have narrower frequency bands. So a reasonable explanation is that the narrower frequency bands help the model concentrates on the specific frequency and not be disturbed by other information. So the model can learn the features of each frequency component and extract domain-invariant features.

About the consistency loss. Then the consistency loss is removed, and significant performance degradation can be observed. The two features for calculating consistency loss represent two complementary frequency components of an image. Since they are supposed to express information about the same object, they should be similar.

The existence of consistency loss helps the framework learn features related to classification tasks, rather than the confusion by interference information related to the frequency domain or source domain. Thus, the classification accuracy is improved in the domain generalization task with the help of consistency loss between the feature extracted from the two branches.

1.3 Sensitivity of Hyper-parameters

The proposed method also introduces some hyper-parameters, and the adjustment of these hyper-parameters is not complicated. We conduct further analysis through the following experiments.

About α . We tried different α values, and the results are shown in Figure 2. In general, the performance of the model is not sensitive to the weight of α . For all tested α , the average performance of the model is always above 60%, which is still significantly higher than the previous methods. A closer observation of the experimental results shows that with the increase of α , the accuracy increases at first. It indicates that urging the network to extract consistent features of different frequencies can improve the effectiveness of features.

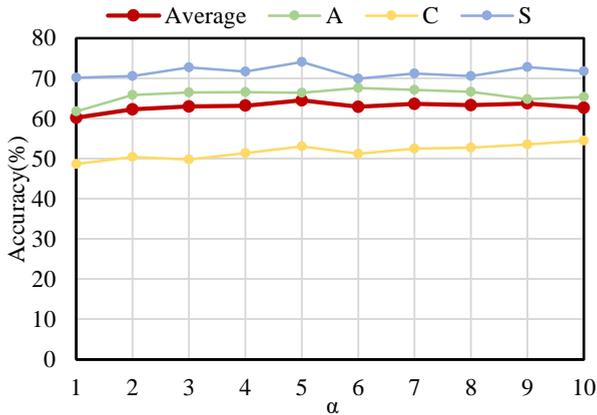


Figure 2: Accuracy of the model on the test domains with the variation of α from 1 to 10. The red line is the average of the other three lines.

Meanwhile, with the further increase of α , the accuracy decreases gradually because the optimization of classification loss is affected. Finally, we selected the hyper-parameter value with the highest accuracy, namely $\alpha = 5$.

About the number of frequency slices K. Another hyper-parameter is related to the division of frequency bands. We tried to decompose the image into a different number of slices and carried out experiments. The results are shown in Table 1.3. All the tested decomposition methods can achieve better performance than previous methods. Among them, decomposing the image into 6 slices is the best. The too rough or fine division will have a certain impact on the performance. The too rough division will make the model unable to fully learn the effective information in each slice, while too fine division will cause too little effective information in each slice, thus increasing the difficulty of learning. For most image classification tasks, we think 6~8 are more appropriate. Furthermore, some automatic partitioning methods could be an improvement direction.

1.4 Visualization of Features

To further demonstrate the effectiveness of our approach, we use t-SNE to visualize the distribution of the unseen target features in the sketch and the art painting domain. We train two models with the ERM method and our approach on the Photo domain and test them on the Sketch and art painting samples separately, in which our approach has the largest and smallest improvement. We use the first 150 samples of each category to the plot.

K	A	C	S	Avg.
2	63.38	44.71	66.00	58.03
4	63.92	54.05	68.77	62.25
6	66.41	53.07	74.10	64.52
8	66.65	48.36	68.94	61.32

Table 2: Accuracy(%) of models trained on original samples and tested on every frequency slide. Low accuracy in each clone proves that many frequencies are not learned well.

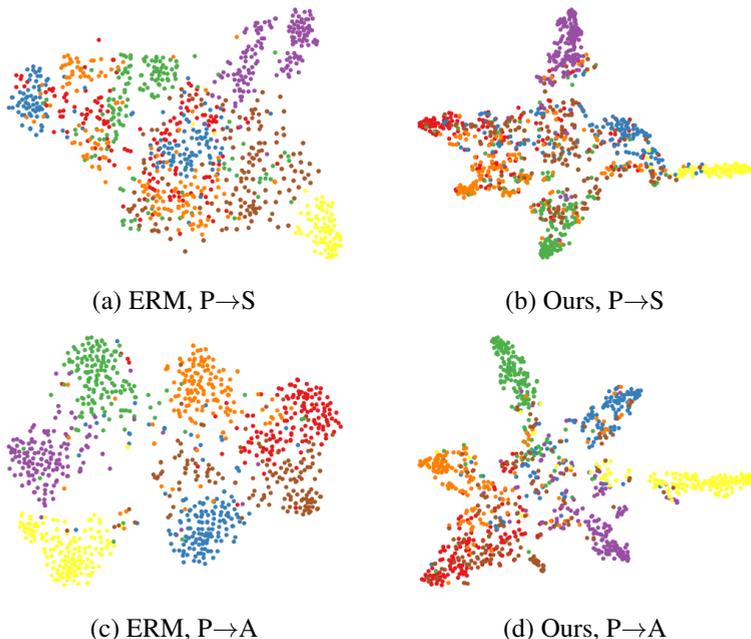


Figure 3: The t-SNE visualizations of target feature distribution for ERM and our approach. The models are trained on Photo and tested on Sketch and Art painting. Features with the same semantic label are drawn in the same color.

It can be seen in Figure 3 that the same category features extracted by our approach are gathered more tightly than by the ERM method. And features of different categories are more distant by our method. Our approach obviously has better class separation than the baseline model, which indicates that our method extracts efficient classification features in different unseen domains.

1.5 Accuracy of Different Frequency Slices

Table 1.4 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

	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

Table 3: 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.

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.

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 4: 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**.

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

Table 5: 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**.

1.6 Similar Information Between Frequency Slices

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 Table 1.5 and Table 1.5, 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.